شبكة الابحاث والدراسات الاقتصادية

الاحصاء الوصفي

أعداد

الدكتور شرف الدين خليل

موقع الشبكة عبر شبكة الانترنيت

www.rr4ee.net

مكتبة

شبكة الابحاث والدراسات الاقتصادية

الفصـــــــل الأول

التعريف بعلم الإحصاء

1/1 مقدمـــــة

من المفاهيم الشائعة بين الناس عن الإحصاء، ما هي إلا أرقام وبيانات رقمية فقط، كأعــداد السكان، وأعداد المواليد، وأعداد الوفيات، وأعداد المزارعين، وأعداد المزارع، وخلافه، ومن ثم ارتبط مفهوم الناس عن الإحصاء بأنه عد أو حصر الأشياء والتعبير عنها بأرقام، وهذا هو المفهوم المحدود لعلم الإحصاء، ولكن الإحصاء كعلم، هو الذي يهتم بطرق جمع البيانات، وتبويبها، وتلخيصها بشكل يمكن الاستفادة منها في وصف البيانات وتحليلها للوصول إلى قرارات سليمة في ظل ظروف عدم التأكد.

2/1 وظائف علم الإحصاء

من التعريف السابق يمكن تحديد أهم وظائف علم الإحصاء في الآتي:

1- وصف البيانات Data Description
2- الاستدلال الإحصائي Statistical Inference
3- التنبؤ Forecasting

أولا: وصف البيانات

تعتبر طريقة جمع البيانات وتبويبها وتلخيصها من أهم وظائف علم الإحــصاء، إذ لا يمكــن الاستفادة من البيانات الخام، ووصف الظواهر المختلفة محل الاهتمام، إلا إذا تم جمع البيانات وعرضها في شكل جدلي، أو بياني من ناحية، وحساب بعض المؤشرات الإحصائية البسيطة التي تدلنا علــى طبيعــة البيانات من ناحية أخرى.

ثانيا: الاستدلال الإحصائي

وهو أيضا من أهم الوظائف المستخدمة في مجال البحث العلمي، ويستند الاستدلال الإحصائي على فكرة اختيار جزء من المجتمع يسمى عينة بطريقة علمية مناسبة، بغرض استخدام بيانات هذه العينة في التوصل إلى نتائج، يمكن تعميمها على مجتمع الدراسة، ومن ثم يهتم الاستدلال الإحصائي بموضوعين هما:

1- التقدير Estimate: وفيه يتم حساب مؤشرات من بيانات العينة تــسمى إحــصاء Statistics تستخدم كتقدير لمؤشرات المجتمع وتسمى معالم Parameters، ويطلق على المقاييس الإحــصائية المحسوبة من بيانات العينة في هذه الحالة بالتقدير بنقطة Point Estimate، كما يمكــن أيــضا استخدام المقاييس الإحصائية المحسوبة من بيانات العينة في تقدير المدى الذي يمكن أن يقع داخلــه معلمة المجتمع باحتمال معين، ويسمى ذلك التقدير بفترة Interval Estimate.

2- اختبارات الفروض Tests of Hypotheses: وفيه يتم استخدام بيانات العينة للوصول إلى قرار علمي سليم بخصوص الفروض المحددة حول معالم المجتمع.

ثالثا: التنبؤ

وفيه يتم استخدام نتائج الاستدلال الإحصائي، والتي تدلنا على سلوك الظاهرة في الماضي في معرفة ما يمكن أن يحدث لها في الحاضر والمستقبل. وهناك العديد من الأساليب الإحصائية المعروفة التي تستخدم في التنبؤ، ومن أبسطها أسلوب الاتجاه العام، وهي معادلة رياضية يتم تقدير معاملاتها باستخدام بيانات العينة، ثم بعد ذلك استخدام المعادلة المقدرة في التنبؤ بما يمكن أن يحدث للظاهرة في المستقبل.

3/1 أنواع البيانات وطرق قياسها

من التعريف السابق لعلم الإحصاء، يلاحظ أنه العلم الذي يهتم بجمع البيانات Data، ونوع البيانات، وطريقة قياسها من أهم الأشياء التي تحدد التحليل الإحصائي المستخدم، وللبيانات أنواع تختلف في طريقة قياسها، ومن الأمثلة على ذلك: بيانات النوع (ذكور Male – إناث Female)، وبيانات تقدير الطالب ($D-D^{+}-C-C^{+}-B-B^{+}-A-A^{+}$)، وبيانات عن درجة الحرارة اللازمة لحفظ الدجاج فترة زمنية معينة، وبيانات عن حجم الإنفاق العائلي بالألف ريال خلال الشهر. ومن هذه الأمثلة نجد أن بيانات النوع غير رقمية، بينما بيانات تقدير الطالب بيانات رقمية موضوعة في شكل مستويات أو فئات، أما بيانات كل من درجة الحرارة، وحجم الإنفاق العائلي فهي بيانات رقمية، ومن ثم يمكن تقسيم البيانات إلى مجموعتين هما:

1- البيانات الوصفية Qualitative Data

2- البيانات الكمية Quantitative Data

أولا: البيانات الوصفية

هي بيانات غير رقمية، أو بيانات رقمية مرتبة في شكل مستويات أو في شكل فئات رقمية، ومن ثم تقاس البيانات الوصفية بمعيارين هما:

أ- بيانات وصفية مقاسة بمعيار اسمي Nominal Scale: وهي بيانات غير رقمية تتكون من مجموعات متنافية، كل مجموعة لها خصائص تميزها عن المجموعة الأخرى، كما أن هذه المجموعات لا يمكن المفاضلة بينها، ومن الأمثلة على ذلك:

- النوع: متغير وصفي تقاس بياناته بمعيار اسمي " ذكر – أنثى " .
- الحالة الاجتماعية: متغير وصفي تقاس بياناته بمعيار اسمي " متزوج ــ أعزب ــ أرمل ــ مطلق ".
- أصناف التمور: متغير وصفي يقاس بياناته بمعيار اسمي " برحي ــ خلاص ــ سكري ــ ....".
- الجنسية: متغير وصفي يقاس بياناته بمعيار اسمي " سعودي ــ غير سعودي"

وهذا النوع من البيانات يمكن تكويد مجموعاته بأرقام، فمثلا الجنسية يمكن إعطاء الجنسية "سعودي" الكود (1)، والجنسية "غير سعودي" الكود (2)

ب- بيانات وصفية مقاسة بمعيار ترتيبي Ordinal Scales: وتتكون من مستويات، أو فئات يمكن ترتيبها تصاعديا أو تنازليا، ومن الأمثلة على ذلك:

- تقدير الطالب: متغير وصفي تقاس بياناته بمعيار ترتيبي "$D-D^{+}-C-C^{+}-B-B^{+}-A-A^{+}$"
- المستوى التعليمي: متغير وصفي تقاس بياناته بمعيار ترتيبي "أمي – يقرأ ويكتب ــ ابتدائية ــ متوسطة ــ ثانوية ــ جامعية ــ أعلى من جامعية "
- تركيز خلات الصوديوم المستخدم في حفظ لحوم الدجاج من البكتريا: متغير وصفي ترتيبي يقاس بياناته بمعيار ترتيبي "0% ــ 5% ــ 10% ــ 15%"
- فئات الدخل العائلي في الشهر بالريال " 5000> ، 5000-10000 ، 10000-15000 ، 15000-20000، 20000< ".

ثانيا: البيانات الكمية

هي بيانات يعبر عنها بأرقام عددية تمثل القيمة الفعلية للظاهرة، وتنقسم إلى قسمين هما:

أ- بيانات فترة Interval Data: وهي بيانات رقمية،تقاس بمقدار بعدها عن الصفر، أي أن للصفر دلالة على وجود الظاهرة، ومن أمثلة ذلك:

- درجة الحرارة: متغير كمي تقاس بياناته بمعيار بعدي، حيث أن درجة الحرارة "0^{o}" ليس معناه انعدام الظاهرة، ولكنه يدل على وجود الظاهرة.
- درجة الطالب في الاختبار: متغير كمي يقاس بياناته بمعيار بعدي، حيث حصول الطالب على الدرجة "0" لا يعني انعدم مستوى الطالب.

ب- بيانات نسبية Ratio Data: هي متغيرات كمية، تدل القيمة "0" على عدم وجود الظاهرة ومن الأمثلة على ذلك:

- إنتاجية الفدان بالطن/هكتار.
- المساحة المتزرعة بالأعلاف بالدونم.
- كمية الألبان التي تنتجها البقرة في اليوم.
- عدد مرات استخدام المزرعة لنوع معين من الأسمدة.
- عدد الوحدات المعيبة من إنتاج المزرعة.

ويلاحظ أن بيانات الفترة لا يمكن إخضاعها للعمليات الحسابية مثل عمليات الضرب والقسمة، بينما يمكن فعل ذلك مع البيانات النسبية.

4/1 طرق جمع البيانات

تعتبر طريقة جمع البيانات من أهم المراحل التي يعتمد عليها البحث الإحصائي، كما أن جمع البيانات بأسلوب علمي صحيح، يترتب عليه الوصول إلى نتائج دقيقة في التحليل،

ولدراسة طرق جمع البيانات، يجب الإلمام بالنقاط التالية:

1- مصادر البيانات. 2- أسلوب جمع البيانات.

3- أنواع العينات 4- وسائل جمع البيانات.

1/4/1 مصادر جمع البيانات

هناك مصدرين للحصول منها على البيانات هما:

1- المصادر الأولية. 2- المصادر الثانوية.

أولا: المصادر الأولية: وهي المصادر التي نحصل منها على البيانات بشكل مباشر، حيث يقوم الباحث نفسه بجمع البيانات من المفردة محل البحث مباشرة، فعندما يهتم الباحث بجمع بيانات عن الأسرة، يقوم بإجراء مقابلة مع رب الأسرة، ويتم الحصول منه مباشرة على بيانات خاصة بأسرته، مثل بيانات المنطقة التابع لها، والحي الذي يسكن فيه، والجنسية، والمهنة، والدخل الشهري، وعدد أفراد الأسرة، والمستوى التعليمي، ... وهكذا.

ويتميز هذا النوع من المصادر بالدقة والثقة في البيانات، لأن الباحث هو الذي يقوم بنفسه بجمع البيانات من المفردة محل البحث مباشرة، ولكن أهم ما يعاب عليها أنها تحتاج إلى وقت ومجهود كبير، ومن ناحية أخرى أنها مكلفة من الناحية المادية.

ثانيا: المصادر الثانوية: وهي المصادر التي نحصل منها على البيانات بشكل غير مباشر، بمعنى آخر يتم الحصول عليها بواسطة أشخاص آخرين، أو أجهزة، وهيئات رسمية متخصصة، مثل نشرات وزارة الزراعة، ونشرات مصلحة الإحصاء، ونشرات منظمة الأغذية " الفاو"....وهكذا.

ومن مزايا هذا النوع من المصادر، توفير الوقت والجهد والمال، إلا أن درجة ثقة الباحث فيها ليست بنفس الدرجة في حالة المصادر الأولية.

2/4/1 أسلوب جمع البيانات

يتحدد الأسلوب المستخدم في جمع البيانات، حسب الهدف من البحث، وحجم المجتمع محل البحث، وهناك أسلوبين لجمع البيانات هما:

1- أسلوب الحصر الشامل. 2- أسلوب المعاينة.

أولا: أسلوب الحصر الشامل: يستخدم هذا الأسلوب إذا كان الغرض من البحث هو حصر جميع مفردات المجتمع، وفي هذه الحالة يتم جمع بيانات عن كل مفردة من مفردات المجتمع بلا استثناء، كحصر جميع المزارع التي تنتج التمور، أو حصر البنوك الزراعية في المملكة، ويتميز أسلوب الحصر الشامل بالشمول وعدم التحيز، ودقة النتائج، ولكن يعاب عليه أنه يحتاج إلى الوقت والمجهود، والتكلفة العالية.

ثانيا: أسلوب المعاينة: يعتم هذا الأسلوب على معاينة جزء من المجتمع محل الدراسة، يتم اختياره بطريقة علمية سليمة، ودراسته ثم تعميم نتائج العينة على المجتمع، ومن ثم يتميز هذا الأسلوب بالآتي:

1- تقليل الوقت والجهد.
2- تقليل التكلفة.
3- الحصول على بيانات أكثر تفصيلا، وخاصة إذا جمعت البيانات من خلال استمارة استبيان.
4- كما أن أسلوب المعاينة يفضل في بعض الحالات التي يصعب فيها إجراء حصر شامل، مثل معاينة دم المريض، أو إجراء تعداد لعدد الأسماك في البحر، أو معاينة اللمبات الكهربائية.

ولكن يعاب على أساوب المعاينة:أن النتائج التي تعتمد على هذا الأسلوب أقل دقة من نتائج أسلوب الحصر الشامل، وخاصة إذا كانت العينة المختارة لا تمثل المجتمع تمثيلا جيدا.

3/4/1 أنواع العينات

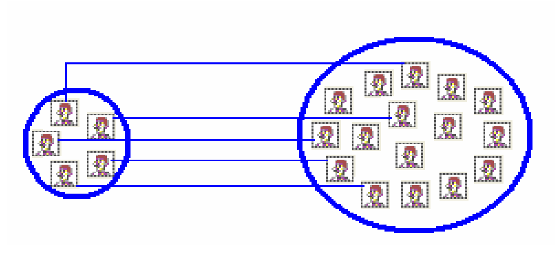
لكي نستعرض أنواع العينات، يتم أولا تحديد الفرق بين مجتمع الدراسة، والعينة المسحوبة من هذا المجتمع.

أ- المجتمع: هو مجموعة من المفردات التي تشترك في صفات، وخصائص محددة، ومجتمع الدراسة هو الذي يشمل جميع مفردات الدراسة، أي هو الكل الذي نرغب دراسته، مثل مجتمع مزارع إنتاج الدواجن، أو مجتمع طلاب الصف الثالث الثانوي.

ب- العينة: هو جزء من المجتمع يتم اختياره بطرق مختلفة بغرض دراسة هذا المجتمع.

شكل رقم (1)

الفرق بين المجتمع والعينة



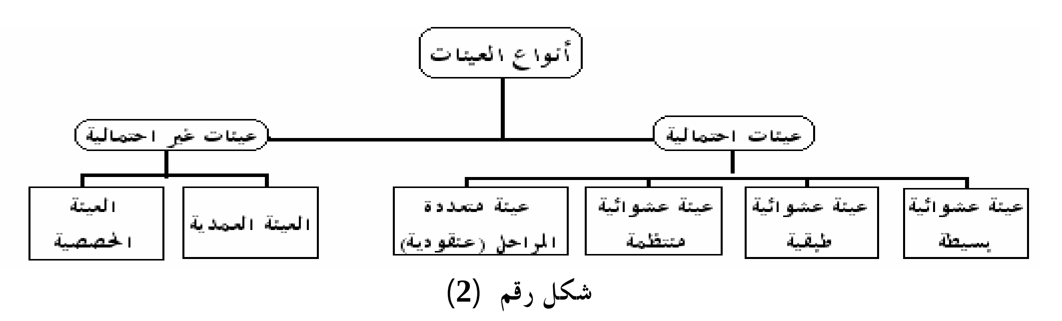
مجتمع الدراسة عينة الدراسة

ويتوقف نجاح استخدام أسلوب المعاينة على عدة عوامل هي:

1- كيفية تحديد حجم العينة. 2- طريقة اختيار مفردات العينة 3- نوع العينة المختارة.

ويمكن تقسيم العينات وفقا لأسلوب اختيارها إلى نوعين هما:

أ- العينات الاحتمالية ب- العينات غير الاحتمالية



شكل رقم (2)

أولا: العينات الاحتمالية

هي العينات التي يتم اختيار مفرداها وفقا لقواعد الاحتمالات، بمعنى آخر هي التي يتم اختيار مفرداها من مجتمع الدراسة بطريقة عشوائية، بهدف تجنب التحيز الناتج عن اختيار المفردات، ومن أهم أنواع العينات الاحتمالية، ما يلي:

أ- العينة العشوائية البسيطة Simple Random Sample.

ب- العينة العشوائية الطبقية Stratified Random Sample.

ت- العينة العشوائية المنتظمة Systematic Random Sample.

ث- العينة العنقودية أو المتعددة المراحل Cluster Sample.

ثانيا: العينات غير الاحتمالية

هي التي يتم اختيار مفرداها بطريقة غير عشوائية، حيث يقوم الباحث باختيار مفردات العينة بالصورة التي تحقق الهدف من المعاينة، مثل اختيار عينة من المزارع التي تنتج التمور من النوع السكري، وأهم أنواع العينات غير الاحتمالية:

أ- العينة العمدية Judgmental Sample

ب- العينة الحصصية Quota Sample

الفصـــــــل الثاني

طرق عرض البيانات

1/2 مقدمـــــــــة

الخطوة التالية بعد جمع البيانات في مجال الإحصاء الوصفي، هو تبويب البيانـــات وعرضـــها بصورة يمكن الاستفادة منها في وصف الظاهرة محل الدراسة، من حيث تمركز البيانات، ودرجة تجانسها. وهناك طريقتين لعرض البيانات هما:

1- عرض البيانات جدوليا.
2- عرض البيانات بيانيا.

2/2 عرض البيانات جدوليا

يمكن عرض البيانات في صورة جدول تكراري، ويختلف شكل الجدول طبقا لنوع البيانـــات، وحسب عدد المتغيرات، وفيما يلي عرض بيانات متغير (وصفي أو كمي) في شكل جدول تكـــراري بسيط.

1/2/2 عرض بيانات المتغير الوصفي في شكل جدول تكراري بسيط

إذا كنا بصدد دراسة ظاهرة ما تحتوي على متغير وصفي واحد، فإنه يمكن عرض بياناتـــه في شكل جدول تكراري بسيط، وهو جدول يتكون من عمودين، أحدهما به مستويات (مجموعات) المتغير، والثاني به عدد المفردات (التكرارات) لكل مستوى (مجموعة).

والمثال التالي يبين لنا كيف يمكن تبويب البيانات الوصفية الخام في شكل جدول تكراري.

مثال (2-1)

فيما يلي بيانات عينة من 40 مزرعة عن نوع التمر الذي تنتجه المزرعة.

سكري	خلاص	برحي	خلاص	برحي	خلاص	صقعي	خلاص
برحي	سكري	برحي	صقعي	خلاص	برحي	نبوت سيف	برحي
صقعي	برحي	سكري	خلاص	برحي	برحي	صقعي	خلاص
برحي	خلاص	برحي	سكري	نبوت سيف	صقعي	نبوت سيف	صقعي
خلاص	برحي	صقعي	نبوت سيف	سكري	برحي	صقعي	خلاص

والمطلوب:

1- ما هو نوع المتغير؟، وما هو المعيار المستخدم في قياس البيانات؟.
2- اعرض البيانات في شكل جدول تكراري.
3- كون التوزيع التكراري النسبي.
4- علق على النتائج.

الحـــل

1- نوع التمر (سكري – خلاص – برحي – صقعي – نبوت سيف) متغير وصفي، تقاس بياناته بمعيار اسمي.

2- لعرض البيانات في شكل جدول تكراري ، يتم إتباع الآتي:

- **تكوين جدول تفريغ البيانات:**

وهو جدول يحتوي على علامات إحصائية، كل علامة تعبر عن تكرار للمجموعة التي ينتمـــي إليها نوع التمر الذي تنتجه المزرعة، وكل خمس علامات تكون حزمة إحصائية، كما هو مبين بالجدول التالي:

جدول تفريغ البيانات

نوع التمر	العلامات الإحصائية	عدد المزارع (التكرارات)
سكري	𝍸	5
خلاص	𝍸 𝍸	10
برحي	𝍸 𝍸 ///	13
صقعي	𝍸 ///	8
نبوت سيف	////	4
Sum		**40**

- **تكوين الجدول التكراري.**

وهو نفس الجدول السابق، باستثناء العود الثاني، ويأخذ الصورة التالية:

جدول رقم **(2-1)**

التوزيع التكراري لعينة حجمها 40 مزرعة حسب نوع التمر الذي تنتجه

نوع التمر	عدد المزارع (التكرارات) (f)	التوزيع التكراري النسبي
سكري	5	$\left(\frac{5}{40}\right) = 0.125$
خلاص	10	$\left(\frac{10}{40}\right) = 0.25$
برحي	13	$\left(\frac{13}{40}\right) = 0.325$
صقعي	8	$\left(\frac{8}{40}\right) = 0.20$
نبوت سيف	4	$\left(\frac{4}{40}\right) = 0.10$
Sum	**40**	**1.00**

المصدر: بيانات افتراضية.

3- التوزيع التكراري النسبي:

يحسب التكرار النسبي بقسمة تكرار المجموعة على مجموع التكرارات، أي أن:

$$\text{التكرار النسبي} = \frac{\text{تكرار المجموعة}}{\text{مجموع التكرارات (n)}} = \left(\frac{f}{\Sigma f}\right) \qquad (٢-١)$$

والعمود الثالث في الجدول رقم (2-1) يعرض التكرار النسبي للمزارعين حسب نوع التمر.

4- التعليق: من الجدول رقم (2-1) يلاحظ أن نسبة المزارع التي تنتج النوع "برحي" في العينة هـي 32.5% وهي أكبر نسبة مما يدل على أن النمط الشائع في إنتاج التمور هو ذلك النوع، بينما نجد أن نسبة المزارع التي تنتج النوع "نبوت سيف" حوالي 10.0% وهي أقل نسبة.

مثال (2-2)

فيما يلي بيانات عن المستوى التعليمي لعينة من 50 فرد.

متوسط	يقرأ ويكتب	ثانوي	متوسط	ثانوي	أعلى من جامعي	متوسط	ابتدائي
يقرا ويكتب	متوسط	ثانوي	ثانوي	متوسط	ثانوي	ابتدائي	متوسط
ابتدائي	ثانوي	يقرا ويكتب	جامعي	ثانوي	ابتدائي	يقرا ويكتب	ثانوي
متوسط	ابتدائي	متوسط	ثانوي	ابتدائي	متوسط	جامعي	متوسط
ثانوي	متوسط	ابتدائي	ثانوي	يقرا ويكتب	ابتدائي	ثانوي	ابتدائي
جامعي	ثانوي	جامعي	ابتدائي	جامعي	أعلى من جامعي	ثانوي	ثانوي
متوسط	يقرا ويكتب						

والمطلوب: **1-** اعرض البيانات في شكل جدول تكراري.

2- كون التوزيع التكراري النسبي، ثم علق على النتائج.

الحـــل

1- عرض البيانات في شكل جدول تكراري:

المستوى التعليمي (يقرأ ويكتب- ابتدائي_ متوسط- ثانوي- جامعي- أعلى من جامعي) متغير وصفي ترتيبي، ويمكن عرض البيانات أعلاه في شكل جدول تكراري بإتباع الآتي:

- تكوين جدول تفريغ البيانات:

جدول تفريغ البيانات

المستوى التعليمي	العلامات الإحصائية	عدد الأفراد (التكرارات)
يقرأ ويكتب	𝍸 |	6
ابتدائي	𝍸 𝍸	10
متوسط	𝍸 𝍸 ||	12
ثانوي	𝍸 𝍸 𝍸	15
جامعي	𝍸	5
أعلى من جامعي	||	2
Sum		**50**

- تكوين الجدول التكراري:

جدول رقم (2-2)

التوزيع التكراري لعينة حجمها 50 فرد حسب المستوى التعليمي

المستوى التعليمي	عدد الأفراد (التكرارات) (f)	التوزيع التكراري النسبي
يقرأ ويكتب	6	0.12
ابتدائي	10	0.20
متوسط	12	0.24
ثانوي	15	0.30
جامعي	5	0.10
أعلى من جامعي	2	0.04
Sum	**50**	**1.00**

المصدر: بيانات عينة

2- تكوين التوزيع التكراري النسبي.

بتطبيق المعادلة رقم (2-1) يمكن حساب التكرارات النسبية، والعمود الثالث في الجدول رقم (2-2) يبن هذا التوزيع،

ومن التوزيع النسبي يلاحظ أن حوالي 30% من أفراد العينة ممن لديهم مؤهل ثانوي، بينما يكون نسبة الأفراد ممن لديهم مؤهل اقل من الثانوي (متوسط، ابتدائي، يقرأ ويكتب) أكثر من 5%، أما نسبة الأفراد الحاصلين على مؤهل أعلى من جامعي حوالي 4% وهي أقل نسبة.

ملاحظات على الجدول

عند تكوين جدول ما لعرض البيانات، يجب مراعاة الآتي:

1- كتابة رقم للجدول.
2- كتابة عنوان للجدول.
3- لكل عمود من أعمدة الجدول عنوان يدل على محتواه.
4- يجب كتابة مصدر البيانات في الجدول.

2/2/2 عرض بيانات المتغير الكمي في شكل جدول تكراري بسيط

بنفس الأسلوب السابق المتبع في تكوين جدول تكراري، يمكن أيضا عرض بيانات المتغير الكمي في شكل جدول تكراري بسيط، ويتكون هذا الجدول من عمودين، الأول يحتوي على فئات تصاعدية للقراءات التي يأخذها المتغير، والثاني يشمل التكرارات أو عدد المفردات التي تنتمي قراءاتها للفئة المناسبة لها، والمثال التالي يبين كيف يمكن عرض البيانات الكمية بيانيا.

مثال (2-3)

فيما يلي بيانات درجات 70 طالب في الاختبار النهائي لمقرر مادة الإحصاء التطبيقي.

56	65	70	65	55	60	66	70	75	56
60	70	61	67	61	71	67	62	71	66
68	72	57	68	72	69	57	71	69	75
72	62	67	73	58	63	66	73	63	65
58	73	74	76	74	80	81	60	74	58
76	82	77	83	77	85	91	78	94	72
79	64	57	79	55	87	64	88	78	62

والمطلوب:

1- كون التوزيع التكراري لدرجات الطلاب.

2- كون التوزيع التكراري النسبي.

3- ما هو نسبة الطلاب الحاصلين على درجة ما بين 70إلى أقل من 80؟

4- ما هو نسبة الطلاب الحاصلين على درجة أقل من 70 درجة؟

5- ما هو نسبة الطلاب الحاصلين على درجة 80 أو أكثر ؟

الحـــل

1- تكوين التوزيع التكراري:

درجة الطالب في الاختبار متغير كمي مستمر، ولكي يتم تبويب البيانات في شـــكل جـــدول تكراري، يتم اتباع الآتي:

- حساب المدى Range(R)

Range = Maximum – Minimum

R = 94 - 55 = 39

- تحديد عدد الفئات Classes(C):

تتحدد عدد الفئات وفقا لاعتبارات منها: رأي الباحث، والهدف من البحث، وحجم البيانات، ويرى كثيرا من الباحثين أن أفضل عدد للفئات يجب أن يتراوح بين 5 إلى 15 ، بفـــرض أن عدد الفئات هو 8 فئات، أي أن: (C=8).

- حساب طول الفئة Length(L):

$$L = \frac{Range}{Classes} = \frac{R}{C} = \frac{39}{8} = 4.875 \approx 5$$

- تحديد الفئات:

الفئة تبدأ بقيمة تسمي الحد الأدنى، وتنتهي بقيمة تسمي الحد الأعلى، ومن ثم نجد أن :

- الحد الأدنى للفئة الأولى هو أقل قراءة (درجة) أي أن الحد الأدنى للفئة الأولى = 55

الحد الأعلى للفئة الأولى = الحد الأدنى + طول الفئة = 55 + L = 55+5=60

إذا الفئة الأولى هي: "55 to les than 60" وتقرأ " من 55 إلى أقل من 60 "

_ الحد الأدنى للفئة الثانية = الحد الأعلى للفئة الأولى = 60

الحد الأعلى للفئة الثانية = الحد الأدنى للفئة + طول الفئة = 5 + 60 = 65

إذا الفئة الثانية هي: "60 to les than 65" وتقرأ "من 60 إلى أقل من 65"

- وبنفس الطريقة يتم تكوين حدود الفئات الأخرى، وهي:

الفئة الثالثة : 65 to les than 70   الفئة الرابعة : 70 to les than 75

الفئة الخامسة: 75 to les than 80   الفئة السادسة: 80 to les than 85

الفئة السابعة: 85 to les than 90   الفئة الثامنة: 90 to les than 95

ويمكن كتابة الفئات بأشكال مختلفة كما هو مبين بجدول تفريغ البيانات:

- تكوين جدول تفريغ البيانات:

جدول تفريغ البيانات

الدرجة			العلامات الإحصائية	عدد الطلاب (التكرارات)
فئات	فئات	فئات		
55 to les than 60	55 – 60	55-	𝍸 𝍸	10
60 to les than 65	60 – 65	60-	𝍸 𝍸 //	12
65 to les than 70	65 – 70	65-	𝍸 𝍸 ///	13
70 to les than 75	70 – 75	70-	𝍸 𝍸 𝍸 /	16
75 to les than 80	75 – 80	75-	𝍸 𝍸	10
80 to les than 85	80 – 85	80-	////	4
85 to les than 90	85 – 90	85-	///	3
90 to les than 95	90 - 95	90-95	//	2
Sum				**70**

- تكوين الجدول التكراري:

جدول رقم (**2-3**)

التوزيع التكراري لعدد 70 طالب حسب درجاتهم في اختبار مقرر الإحصاء

فئات الدرجة	عدد الطلاب (التكرارات) (f)	التكرار النسبي
55 – 60	10	0.143
60 – 65	12	0.171
65 – 70	13	0.186
70 – 75	16	0.229
75 – 80	10	0.143
80 – 85	4	0.057
85 – 90	3	0.043
90 – 95	2	0.028
Sum	**70**	**1.00**

المصدر: بيانات نتيجة العام **1426هـ**

2- التوزيع التكراري النسبي:

$$\text{التكرار النسبي} = \frac{f}{n}$$

والعمود الثالث في الجدول رقم (**2-3**) يبين التكرار النسبي.

3- نسبة الطلاب الحاصلين على درجات ما بين 70 إلى أقل من 80 هو مجموع التكرارين النسبيين للفئتين الرابعة والخامسة:

$0.229+0.143=0.372$ = نسبة الطلاب الحاصلين على درجات ما بين (70 , 80)

أي حوالي %37.2 من الطلاب حصلوا على درجات ما بين (70 , 80) .

4- نسبة الطلاب الحاصلين على درجات أقل من 70، هو مجموع التكرارات النسبية للفئات الأولى والثانية، والثالثة:

$0.143+0.171+0.186=0.5$ = نسبة الطلاب الحاصلين على درجة أقل من 70

أي أن حوالي %50 من الطلاب حصلو على درجة أقل من 70 درجة

5- نسبة الطلاب الحاصلين على درجة 80 أو أكثر، هو مجموع التكرارات النسبية للفئات الثلاث الأخيرة:

$0.057+0.043+0.028=0.128$ = نسبة الطلاب الحاصلين على درجات 80أو أكثر

أي أن حوالي %12.8 من الطلاب حصلوا على درجة 80 أو أكثر.

3/2 العرض البياني للبيانات الكمية

العرض البياني للبيانات، هو أحد طرق التي يمكن استخدامها في وصف البيانات، من حيث شكل التوزيع ومدى تمركز البيانات، وفي كثير من النواحي التطبيقية يكون العرض البياني أسهل وأسرع في وصف الظاهرة محل الدراسة، وتختلف طرق عرض البيانات بيانيا حسب نوع البيانات المبوبة في شكل جدول تكراري، وفيما يلي عرض للأشكال البيانية المختلفة.

1/3/2 المدرج التكراري Histogram

المدرج التكراري هو التمثيل البياني للجدول التكراري البسيط الخاص بالبيانات الكمية المتصلة، وهو عبارة عن أعمدة بيانية متلاصقة، حيث تمثل التكرارات على المحور الرأسي، بينما تمثل قيم المتغير (حدود الفئات) على المحور الأفقي، ويتم تمثيل كل فئة بعمود، ارتفاعه هو تكرار الفئة، وطول قاعدته هو طول الفئة.

مثال (2-4)

فيما يلي التوزيع التكراري لأوزان عينة من الدواجن بالجرام، حجمها 100 اختيرت من أحد المزارع بعد 45 يوم.

الوزن	600-	620-	640-	660-	680-	700-720	Sum
عدد الدجاج	10	15	20	25	20	10	100

والمطلوب:

1- ما هو طول الفئة؟

2- ارسم المدرج التكراري.

3- ارسم المدرج التكراري النسبي، ثم علق على الرسم.

الحـــل

1- طول الفئة (L)

طول الفئة = الحد الأعلى للفئة – الحد الأدنى للفئة

$$L = upper - Lower \quad (٢-٢)$$

$$L = 620 - 600 = 640 - 620 = \ldots = 720 - 700 = 20$$

إذا طول الفئة = 20

2- رسم المدرج التكراري.

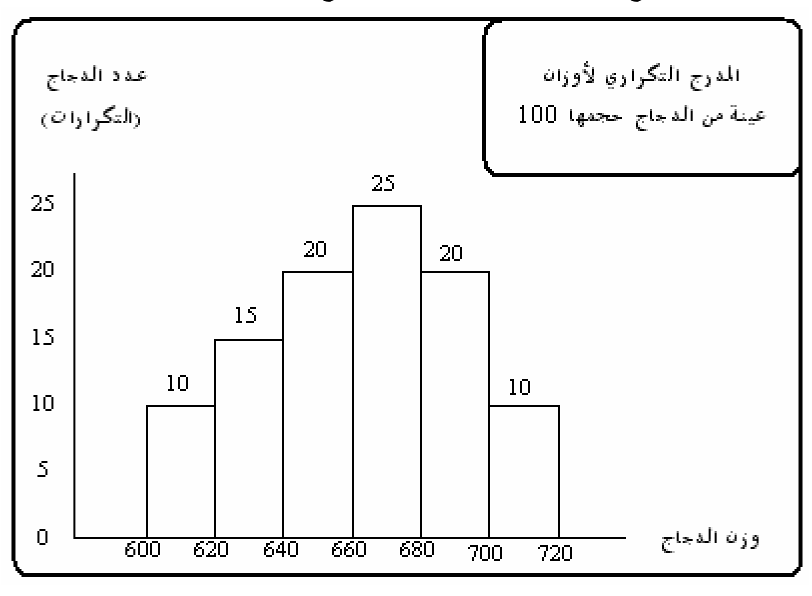
لرسم المدرج التكراري يتم إتباع الخطوات التالية:

- رسم محوران متعامدان، الرأسي ويمثل التكرارات، الأفقي ويمثل الأوزان.
- كل فئة تمثل بعمود ارتفاعه هو تكرار الفئة، وطول قاعدته هو طول الفئة.
- كل عمود يبدأ من حيث انتهى به عمود الفئة السابقة.

والشكل (2-1) يبين المدرج التكراري لأوزان الدجاج.

شكل (2-1)

المدرج التكراري لأوزان عينة من الدجاج حجمها 100 دجاجة



3- رسم المدرج التكراري النسبي: لرسم المدرج التكراري النسبي يتم إجراء الآتي:

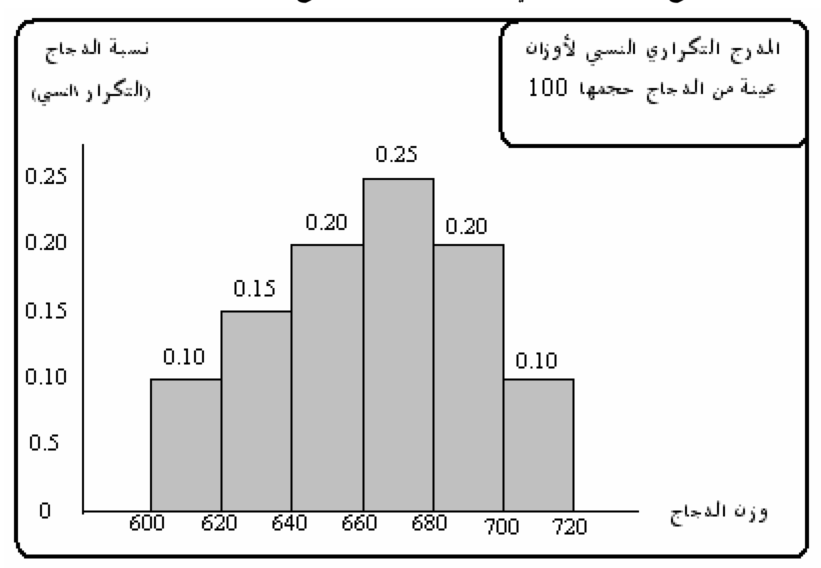
- حساب التكرارات النسبية.

الوزن	600-	620-	640-	660-	680-	700-720	Sum
عدد الدجاج	10	15	20	25	20	10	100
التكرار النسبي	0.10	0.15	0.20	0.25	0.20	0.10	1.00

- بإتباع نفس الخطوات السابقة عند رسم المدرج التكراري، يتم رسم المدرج التكراري النسبي، بإحلال التكرارات النسبية محل التكرارات المطلقة على المحور الرأسي، كما هو مبين في الشكل التالي:

شكل (2-2)

المدرج التكراري النسبي لأوزان عينة من الدجاج حجمها 100 دجاجة



ومن الشكل أعلاه يلاحظ الآتي:

- أن 25% من الدجاج يتراوح وزنه بين 660 ، 680 جرام وهي أكبر نسبة.
- أن الشكل ملتوي جهة اليسار، مما يدل على أن توزيع أوزان الدجاج سالب الالتواء.

ملاحظات على شكل المدرج التكراري

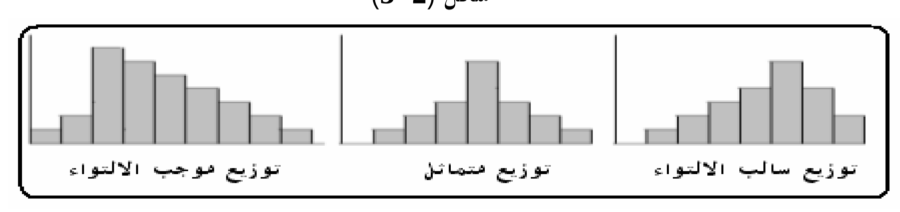
أ- أن المساحة أسفل المدرج التكراري تساوي مجموع التكرارات (n).

ب- أما المساحة أسفل المدرج التكراري النسبي، فهي تعبر عن مجموع التكرارات النسبية، وهي تساوي الواحد الصحيح.

ت- يمكن تقدير القيم الشائعة، وهي القيم التي يناظرها أكبر ارتفاع، ففي الشكلين السابقين، نجد أن الوزن الشائع يقع في الفئة (680-660) ويطلق عليه المنوال.

ث- يمكن معرفة شكل توزيع البيانات، كما هو مبين بالأشكال الثلاث التالية:

شكل (2-3)



2/3/2 المضلع التكراري

هو تمثيل بياني أيضا للجدول التكراري البسيط، حيث تمثل التكرارات على المحور الرأسـي، ومراكز الفئات على المحور الأفقي، ثم التوصيل بين الإحداثيات بخطوط منكسرة، وبعد ذلك يتم توصيل طرفي المضلع بالمحور الأفقي.

ومركز الفئة هي القيمة التي تقع في منتصف الفئة، وتحسب بتطبيق المعادلة التالية:

$$\text{مركز الفئة} = \frac{\text{الحد الأدنى للفئة} + \text{الحد الأعلى للفئة}}{٢} \qquad (٢-٣)$$

$$Midpoint = \frac{Lower + Upper}{2}$$

ونظرا لعدم معرفة القيم الفعلية لتكرار كل فئة، يعتبر مركز الفئة هو التقدير المناسب لقيمـة كل مفردة من مفردات الفئة.

مثال (2-5)

استخدم بيانات الجدول التكراري في المثال (2-4) لرسم المضلع التكراري.

الحـــل

لرسم المضلع التكراري يتبع الآتي:

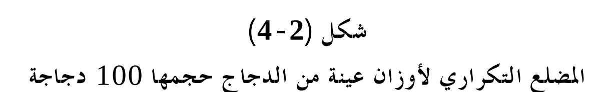
- حساب مراكز الفئات بتطبيق المعادلة رقم (2-3)

الوزن	عدد الدجاج (التكرار)	مركز الفئة (x)
600-	10	(600+620)/2= 610
620-	15	(620+640)/2=630
640-	20	650
660-	25	670
680-	20	690
700-720	10	(700+720)/710
Sum	100	

- نقط الإحداثيات هي :

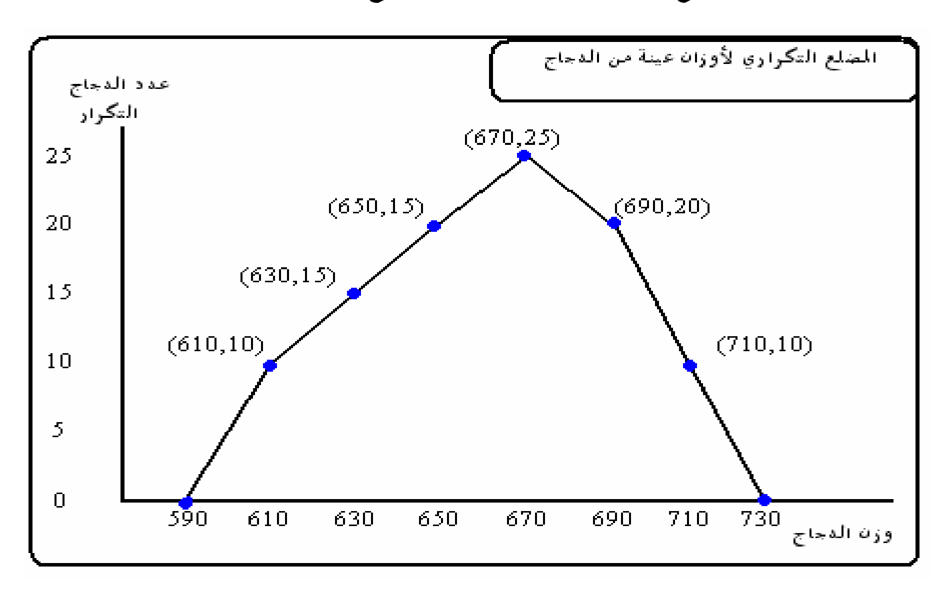
مركز الفئة (x)	590	610	630	650	670	690	710	730
التكرار (y)	0	10	15	20	25	20	10	0

- التمثيل البياني لنقط الإحداثيات وتوصيلها بخطوط مستقيمة، كما هو مبين بالشكل (2-4)



شكل (2-4)

المضلع التكراري لأوزان عينة من الدجاج حجمها 100 دجاجة

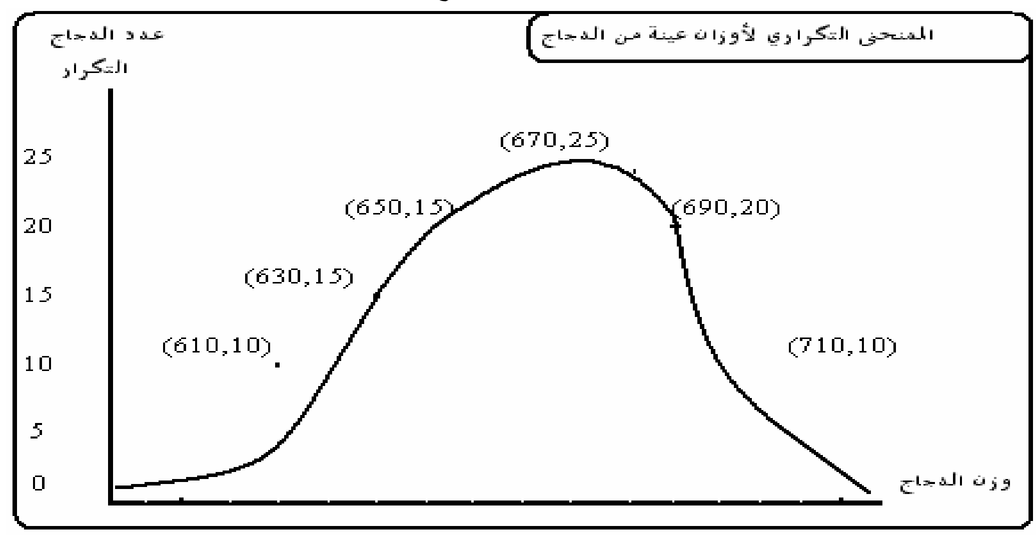


3/3/2 المنحنى التكراري

بإتباع نفس الخطوات السابقة في رسم المضلع يمكن رسم المنحنى التكراري، ولكن يتم تمهيد الخطوط المنكسرة في شكل منحنى بحيث يمر بأكثر عدد من النقاط، وفي المثال السابق يمكن رسم المنحنى التكراري، والشكل (2-5) يبين هذا الشكل.

شكل (2-5)

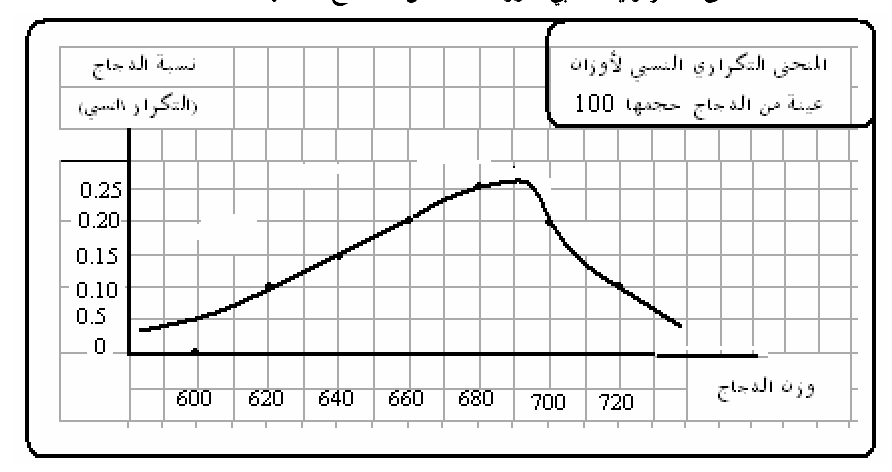
المنحنى التكراري لأوزان عينة من الدجاج حجمها 100 دجاجة



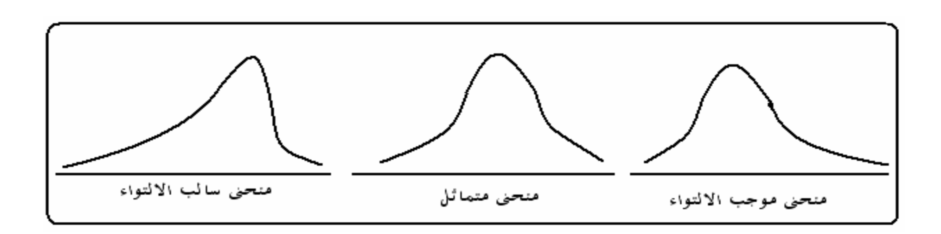
كما يمكن رسم المنحنى التكراري النسبي بتمثيل التكرارات النسبية على المحور الرأسي بدلا من التكرارات المطلقة، ومن ثم يأخذ هذا المنحنى الشكل رقم (2-6) التالي:

شكل (2-6)

المنحنى التكراري النسبي لأوزان عينة من الدجاج حجمها 100 دجاجة



والمنحنى التكراري أعلاه موجب الالتواء، كما أن المساحة أسفل هذا المنحنى تعبر عن مجموع التكرارات النسبية، أي أنها تساوي الواحد الصحيح، وهناك أشكال مختلفة للمنحنى التكراري النسبي، تدل على أشكال توزيع البيانات، ومن أهمها ما يلي:



3/3 التوزيعات التكرارية المتجمعة

في كثير من الأحيان قد يحتاج الباحث إلى معرفة عدد المشاهدات التي تقل عن قيمة معينة أو تزيد عن قيمة معينة، ومن ثم يلجأ الباحث إلى تكوين جداول تجميعية صاعدة أو هابطة، وفيما يلي بيان كيفية تكوين كل نوع من هذين النوعين على حدة:

1/3/3 التوزيع التكراري المتجمع الصاعد

لتكوين الجدول التكراري المتجمع الصاعد، يتم حساب مجموع التكرارات (عدد القيم) التي تقل عن كل حد من حدود الفئات.

مثال (2-6)

الجدول التكراري التالي يبين توزيع 40 بقرة في مزرعة حسب كمية الألبان التي تنتجها البقرة في اليوم باللتر.

كمية الألبان	18-	22-	26-	30-	34-38	Sum
عدد الأبقار	4	9	15	8	4	40

والمطلوب:

1- كون جدول التوزيع التكراري المتجمع الصاعد.
2- كون جدول التوزيع التكراري المتجمع الصاعد النسبي.
3- ارسم المنحنى التكراري المتجمع الصاعد النسبي.
4- من المنحنى المتجمع أوجد الآتي:
- نسبة الأبقار التي يقل إنتاجها عن 28 لتر.
- كمية الإنتاج التي يقل عنها 25% من الأبقار.
- كمية الإنتاج التي يقل عنها 50% من الإنتاج.

الحل

1- التوزيع التكراري المتجمع الصاعد.

توزيع تكراري متجمع صاعد

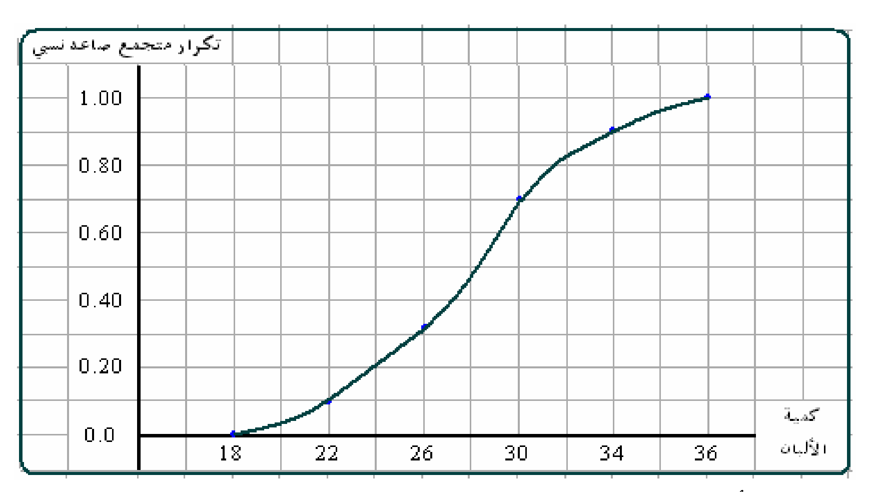
تكرار متجمع صاعد نسبي	تكرار متجمع صاعد	أقل من
0.00	0	أقل من 18
0.10	4	أقل من 22
0.325	13	أقل من 26
0.70	28	أقل من 30
0.90	36	أقل من 34
1.00	40	أقل من 38

التوزيع التكراري

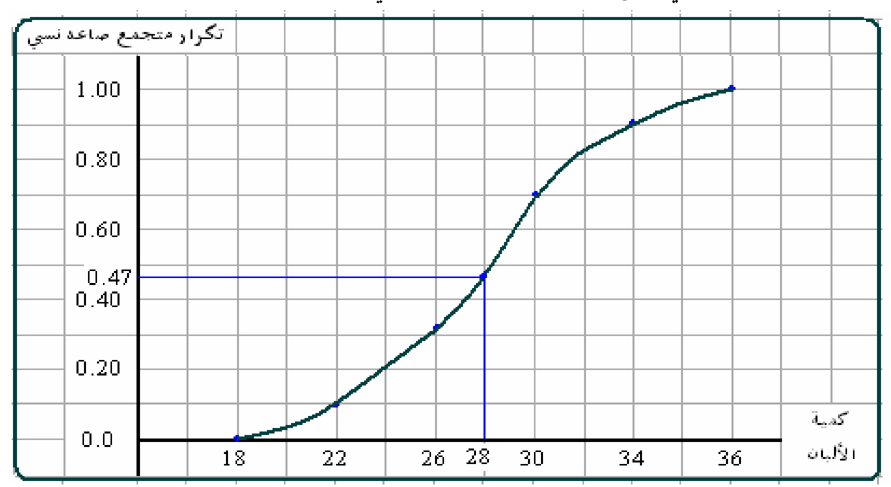
عدد الأبقار	كمية الإنتاج باللتر
4	18-
9	22-
15	26-
8	30-
4	34-38
40	Sum

2- التوزيع التكراري المتجمع الصاعد النسبي: يحسب التكرار المتجمع الصاعد النسبي بقسمة التكرار المتجمع الصاعد على مجموع التكرارات، كما هو مبين بالعمود الأخير في جدول التوزيع التكراري المتجمع الصاعد.

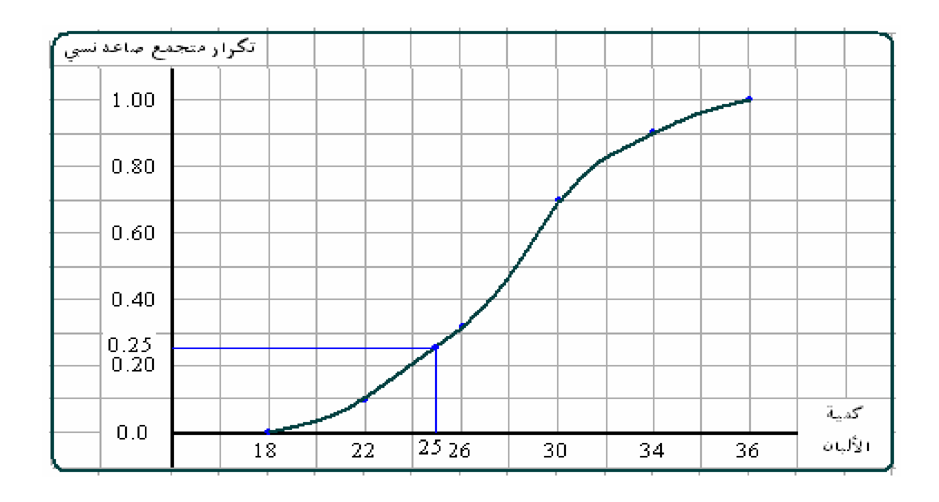
3- رسم المنحنى التكراري المتجمع الصاعد: المنحنى التكراري المتجمع الصاعد النسبي هو التمثيل البياني للتوزيع التكراري المتجمع الصاعد النسبي، حيث تمثل حدود الفئات على المحور الأفقي، والتكرار المتجمع الصاعد النسبي على المحور الرأسي، ويتم تمهيد المنحنى ليمر بالإحداثيات، كما هو مبين في الشكل التالي:



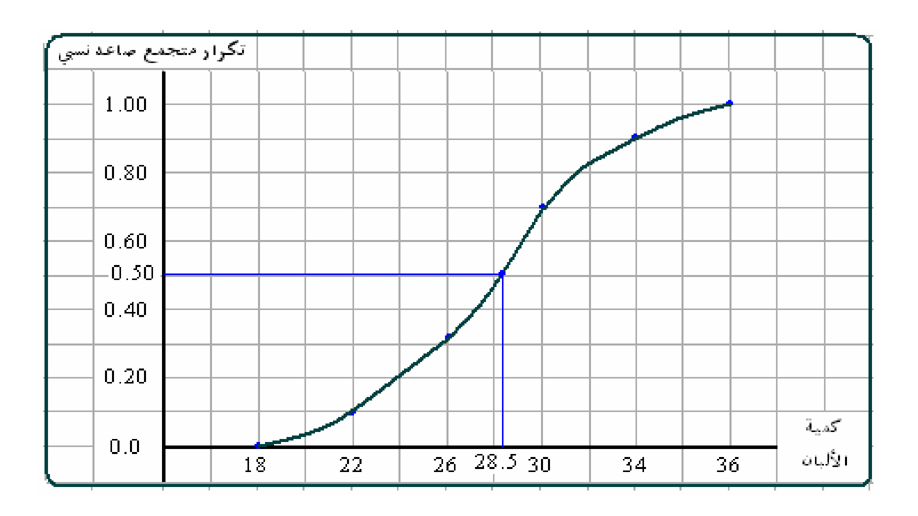
- نسبة الأبقار التي يقل إنتاجها عن 28 لتر هي 0.47 تقريبا.



- كمية الإنتاج التي يقل عنها 25% من قيم الإنتاج هي: 25 لتر تقريبا.



- كمية الإنتاج التي يقل عنها 50% من قيم الإنتاج هي: 28.5 لتر، ويطلق عليها الوسيط:



2/3/3 التوزيع التكراري المتجمع الهابط (النازل)

لتكوين الجدول التكراري المتجمع النازل، يتم حساب مجموع التكرارات (عدد القيم) التي تساوي أو تزيد عن كل حد من حدود الفئات.

مثال (2-7)

استخدم بيانات الجدول التكراري في مثال (2-6)، وأوجد الآتي:

1- كون التوزيع التكراري المتجمع النازل.

2- ارسم المنحنى التكراري المتجمع النازل النسبي.

الحل:

1- تكوين التوزيع التكراري المتجمع النازل.

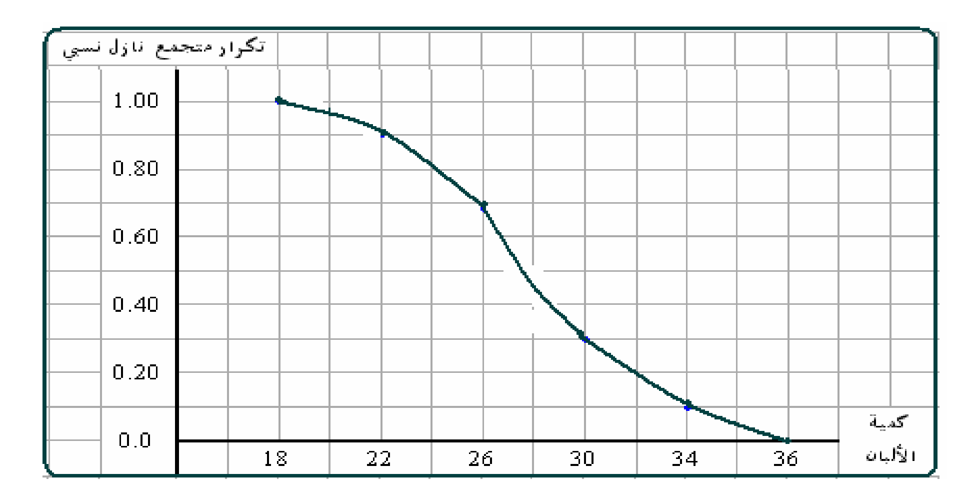
توزيع تكراري متجمع نازل

تكرار متجمع نازل نسبي	تكرار متجمع نازل	أكثر من أو يساوي
1.00	40	أكثر من أو يساوي 18
0.90	36	أكثر من أو يساوي 22
0.675	27	أكثر من أو يساوي 26
0.30	12	أكثر من أو يساوي 30
0.10	4	أكثر من أو يساوي 34
0.00	0	أكثر من أو يساوي 38

التوزيع التكراري

عدد الأبقار	كمية الإنتاج باللتر
4	18-
9	22-
15	26-
8	30-
4	34-38
40	Sum

رسم المنحنى التكراري المتجمع النازل.



ملاحظات:

1- يمكن رسم المنحنيان في شكل بياني واحد، ويلاحظ أنهما يتقاطعان عند نقطة تسمى الوسيط.

2- يكون استخدامنا للمنحنى المتجمع الصاعد أكثر وأوقع من الناحية التطبيقية.

4/3 العرض البياني للبيانات الوصفية

يمكن عرض البيانات الخاصة بمتغير وصفي في شكل دائرة بيانية أو أعمدة بيانية، يمكن من خلاله وصف ومقارنة مجموعات أو مستويات هذا المتغير.

1/4/3 الدائرة البيانية

لعرض بيانات المتغير الوصفي في شكل دائرة، يتم توزيع الـ 360^o درجة حسب التكرار النسبي لمجموعات المتغير، حيث تحدد مقدار الزاوية الخاصة بالمجموعة رقم r بتطبيق المعادلة التالية:

التكرار النسبي للمجموعة × 360^o = مقدار الزاوية

مثال (2-8)

الجدول التكراري التالي يبين توزيع عينة حجمها 500 أسرة حسب المنطقة التي تنتمي إليها.

المنطقة	الرياض	الشرقية	القصيم	الغربية	sum
عدد الأسر	150	130	50	170	500

مثل البيانات أعلاه في شكل دائرة بيانية.

الحل:

1- تحديد مقدار الزاوية المخصصة لكل منطقة، بتطبيق المعادلة:

التكرار النسبي للمنطقة × 360^o = مقدار الزاوية المخصص للمنطقة

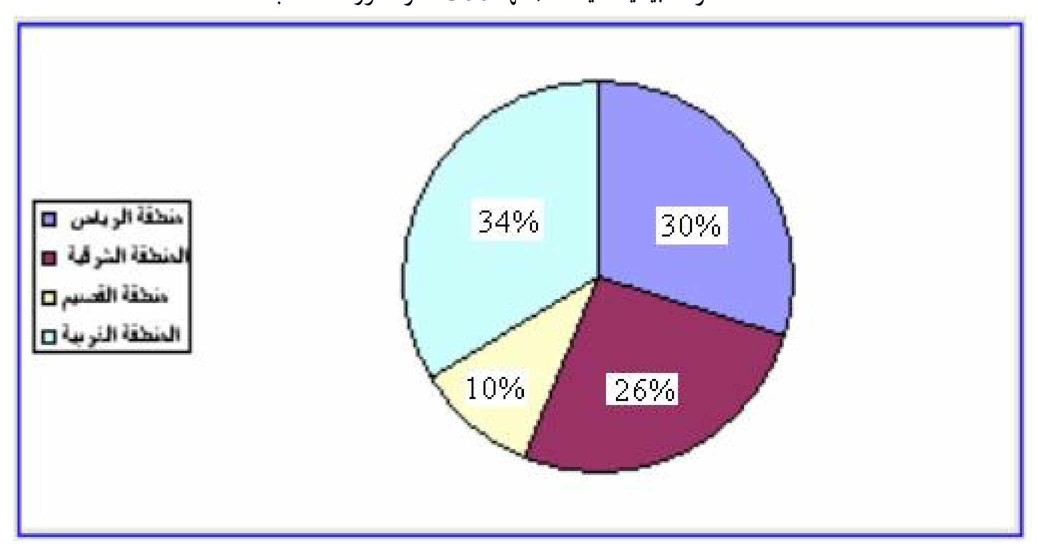
المنطقة	عدد الأسر	التكرار النسبي	مقدار الزاوية
الرياض	150	0.30	$360 \times 0.30 = 108^o$
الشرقية	130	0.26	$360 \times 0.26 = 93.6^o$
القصيم	50	0.10	$360 \times 0.10 = 36^o$
الغربية	170	0.34	$360 \times 0.30 = 122.4^o$
Sum	500	1.00	360^o

2- رسم الدائرة

يتم رسم دائرة وتقسيمها إلى أربع أجزاء لكل منطقة جزء يتناسب مع مقدار الزاوية المخصصة له، كما هو مبين في الشكل التالي:

شكل رقم (2-7)

الدائرة البيانية لعينة حجمها 500 أسرة موزعة حسب المنطقة



ومن الشكل أعلاه يلاحظ أن نسبة الأسر التي تنتمي للمنطقة الغربية حوالي 34% وهي أكبر نسبة في العينة، بينما يكون نسبة الأسر في منطقة القصيم حوالي 10% وهي أقل نسبة في العينة.

الفصـــــــل الثالث

مقاييس النزعة المركزية

Central Tendency

1/3 مقدمة

في كثير من النواحي التطبيقية يكون الباحث في حاجة إلى حساب بعض المؤشرات التي يمكن الاعتماد عليها في وصف الظاهرة من حيث القيمة التي تتوسط القيم أو تنزع إليها القيم ، ومن حيث التعرف على مدى تجانس القيم التي يأخذها المتغير، وأيضا ما إذا كان هناك قيم شاذة أم لا . والاعتماد على العرض البياني وحدة لا يكفى ، ولذا يتناول هذا الفصل، والذي يليه عرض بعض المقاييس الإحصائية التي يمكن من خلالها التعرف على خصائص الظاهرة محل البحث، وكذلك إمكانية مقارنة ظاهرتين أو أكثر ، ومن أهم هذه المقاييس ، مقاييس النزعة المركزية والتشتت .

2/3 مقاييس النزعة المركزية

تسمى مقاييس النزعة المركزية بمقاييس الموضع أو المتوسطات ، وهى القيم التى تتركز القيم حولها ، ومن هذه المقاييس ، الوسط الحسابي ، والمنوال ، والوسيط ، والوسط الهندسي ، والوسط التوافقي ، والرباعيات ، والمئينات ، وفيما يلي عرض لأهم هذه المقاييس

1/2/3 الوسط الحسابي Arithmetic Mean

من أهم مقاييس النزعة المركزية ، وأكثرها استخداما في النواحي التطبيقية ، ويمكن حسابه للبيانات المبوبة وغير المبوبة ، كما يلي :

أولا: الوسط الحسابي للبيانات غير المبوبة

يعرف الوسط الحسابي بشكل عام على أنه مجموع القيم مقسوما على عددها . فإذا كان لدينا n من القيم ، ويرمز لها بالرمز : $x_1, x_2, \ldots, x_n$.

فإن الوسط الحسابي لهذه القيم ، ونرمز له بالرمز $\overline{x}$ يحسب بالمعادلة التالية :

$$\text{الوسط الحسابي} = \frac{\text{مجموع القيم}}{\text{عدد القيم}} \qquad (1-3)$$

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

حيث يدل الرمز Σ على المجموع .

مثـــال (3-1)

فيما يلي درجات 8 طلاب في مقرر **122** إحصاء تطبيقي .

34   32   42   37   35   40   36   40

والمطلوب إيجاد الوسط الحسابي لدرجة الطالب في الامتحان .

الحـــل

لإيجاد الوسط الحسابي للدرجات تطبق المعادلة رقم (**3-1**) كما يلي:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

$$= \frac{34+32+42+37+35+40+36+40}{8} = \frac{296}{8} = 37$$

أي أن الوسط الحسابي لدرجة الطالب في اختبار مقرر **122** إحص يساوي 37 درجة

ثانيا: الوسط الحسابي للبيانات المبوبة

من المعلوم أن القيم الأصلية ، لا يمكن معرفتها من جدول التوزيع التكراري ، حيث أن هذه القيم موضوعة في شكل فئات ، ولذا يتم التعبير عن كل قيمة من القيم التي تقع داخل حدود الفئة بمركز هذه الفئة ، ومن ثم يؤخذ في الاعتبار أن مركز الفئة هو القيمة التقديرية لكل مفردة تقع في هذه الفئة.

فإذا كانت k هي عدد الفئات ، وكانت $x_1, x_2, \ldots, x_k$ هي مراكز هذه الفئات، $f_1, f_2, \ldots, f_k$ هي التكرارات ، فإن الوسط الحسابي يحسب بالمعادلة التالية:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_k f_k}{f_1 + f_2 + \ldots + f_k} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i} \qquad (٣-٢)$$

مثـــال (**3-2**)

الجدول التالي يعرض توزيع 40 تلميذ حسب أوزانهم .

فئات الوزن	32-34	34-36	36-38	38-40	40-42	42-44
عدد التلاميذ	4	7	13	10	5	1

والمطلوب إيجاد الوسط الحسابي.

الحـــــل

لحساب الوسط الحسابي باستخدام المعادلة رقم (3-2) يتم إتباع الخطوات التالية :

1- إيجاد مجموع التكرارات $\sum f$. **2-** حساب مراكز الفئات x .

3- ضرب مركز الفئة في التكرار المناظر له (xf)، وحساب المجموع $\sum xf$

4- حساب الوسط الحسابي بتطبيق المعادلة رقم (3-2) .

فئات الوزن (C)	التكرارات f	مراكز الفئات x	$x\ f$
32-34	4	2=33÷(32+34)	33=132×4
34-36	7	35	35=245×7
36-38	13	37	37=481×13
38-40	10	39	39=390×10
40-42	5	41	41=205×5
42-44	1	43	43=43×1
المجموع	40		1496

إذا الوسط الحسابي لوزن التلميذ هو :

$$\bar{x} = \frac{\sum_{i=1}^{6} x_i f_i}{\sum_{i=1}^{6} f_i} = \frac{1496}{40} = 37.4\ k.g$$

أي أن متوسط وزن التلميذ يساوي 37.4 k.g

خصائص الوسط الحسابي

يتصف الوسط الحسابي بعدد من الخصائص ، ومن هذه الخصائص ما يلي :

1- الوسط الحسابي للمقدار الثابت يساوى الثابت نفسه ، أي أنه إذا كانت قيم x هي : $x : a , a ,..., a$ ، فإن الوسط الحسابي هو:

$$\bar{x} = \frac{a + a + \ldots + a}{n} = \frac{na}{n} = a \qquad (٣-٣)$$

ومثال على ذلك ، لو اخترنا مجموعة من 5 طلاب ، ووجدنا أن كل طالب وزنه 63 كيلوجرام ، فإن متوسط وزن الطالب في هذه المجموعة هو :

$$\bar{x} = \frac{63+63+63+63+63}{5} = \frac{315}{5} = 63\ k.g$$

2- مجموع انحرافات القيم عن وسطها الحسابي يساوى صفرا ، ويعبر عن هذه الخاصية بالمعادلة .

$$\sum (x - \bar{x}) = 0 \qquad (٤-٣)$$

ويمكن التحقق من هذه الخاصية باستخدام بيانات مثال (3-1) ، نجد أن درجات الطلاب هي

34, 32, 42, 37, 35, 40, 36, 40: ، والوسط الحسابي للدرجة هو $\bar{x}=37$ ، إذا :

x	34	32	42	37	35	40	36	40	296
$(x-\bar{x})$	34-37	32-37	42-37	37-37	35-37	40-37	36-37	40-37	0
$(x-37)$	-3	-5	5	0	-2	3	-1	3	

أي أن : $\sum(x-37)=0$

3- إذا أضيف مقدار ثابت إلى كل قيمة من القيم ، فإن الوسط الحسابي للقيم المعدلة (بعد الإضافة) يساوى الوسط الحسابي للقيم الأصلية (قبل الإضافة) مضافا إليها هذا المقدار الثابت . فإذا كانت القيم هي : $x_1, x_2, \ldots, x_n$ ، وتم إضافة مقدار ثابت (a) إلى كل قيمة من القيم ، ونرمز للقيم الجديدة بالرمز y ، أي أن $y=x+a$ ، فإن : الوسط الحسابي لقيم y (القيم بعد الإضافة) هو:

$$\bar{y} = \bar{x} + a \qquad (٣-٥)$$

حيث أن $\bar{y}$ هو الوسط الحسابي للقيم الجديدة ، ويمكن التحقق من هذه الخاصية باستخدام بيانات مثال رقم **(3-1)** .

إذا قرر المصحح إضافة 5 درجات لكل طالب ، فإن الوسط الحسابي للدرجات المعدلة يصبح قيمته {(37+5)=42} ، والجدول التالي يبين ذلك .

x	34	32	42	37	35	40	36	40	296
$y=(x+5)$	34+5	32+5	42+5	37+5	35+5	40+5	36+5	40+5	336
	39	37	47	42	40	45	41	45	

نجد أن مجموع القيم الجديدة هو : $\sum y = 336$ ، ومن ثم يكون الوسط الحسابي للقيم الجديدة هو :

$$\bar{y} = \frac{\sum y}{n} = \frac{336}{8} = 42 \longrightarrow (\bar{x}+5 = 37+5 = 42)$$

4- إذا ضرب مقدار ثابت (a) في كل قيمة من القيم ، فإن الوسط الحسابي للقيم المعدلة (القيم الناتجة بعد الضرب) يساوى الوسط الحسابي للقيم الأصلية (القيم بعد التعديل) مضروبا في هذا المقدار الثابت . أى أنه إذا كان : $y=a\ x$ ، ويكون الوسط الحسابي للقيم الجديدة y هو :

$$\bar{y} = a\ \bar{x} \qquad (٣-٦)$$

ويمكن للطالب أن يتحقق من هذه الخاصية باستخدام نفس بيانات المثال السابق . فإذا كان تصحيح الدرجة من 50 ، وقرر المصحح أن يجعل التصحيح من 100 درجة ، بمعنى أنه سوف يضرب كل درجة في قيمة ثابتة (a=2) ، ويصبح الوسط الحسابي الجديد هو :

$$\bar{y}=a\ \bar{x}=2(37)=74$$

5- مجموع مربعات انحرافات القيم عن وسطها الحسابي أقل ما يمكن ، أي أن:

$$\sum (x-\bar{x})^2 \langle \sum (x-a)^2 \quad if \quad a \neq \bar{x} \qquad (٧-٣)$$

وفي المثال السابق فإن : $\sum (x-37)^2 < \sum (x-a)^2$ لجميع قيم $a \neq 37$

ثالثا: الوسط الحسابي المرجح

في بعض الأحيان يكون لكل قيمة من قيم المتغير أهمية نسبية تسمى أوزن ، أو ترجيحات ، وعدم أخذ هذه الأوزان في الاعتبار عند حساب الوسط الحسابي ، تكون القيمة المعبرة عن الوسط الحسابي غير دقيقة ، فمثلا لو أخذنا خمسة طلاب ، وسجلنا درجات هؤلاء الطلاب في مقرر الإحصاء التطبيقي ، وعدد ساعات الاستذكار في الأسبوع .

مسلسل	1	2	3	4	5	sum
x (الدرجة)	23	40	36	28	46	173
w (عدد ساعات الاستذكار)	1	3	3	2	4	

نجد أن الوسط الحسابي غير المرجح للدرجة الحاصل عليها الطالب هي :

$$\bar{x}=\frac{\sum x}{n}=\frac{23+40+36+28+46}{5}=\frac{173}{5}=346$$

وإذا أردنا أن نحسب الوسط الحسابي للدرجات x المرجحة بعدد ساعات الاستذكار w ، يتم تطبيق المعادلة التالية :

$$(\bar{w})=\frac{\sum xw}{\sum w}=\frac{23\times1+40\times3+36\times3+28\times2+46\times4}{1+3+3+2+4}$$

$$=\frac{23+120+108+56+184}{13}=\frac{491}{13}=37.769$$

وهذا الوسط المرجح أكثر دقة من الوسط الحسابي غير المرجح .

إذا الوسط الحسابي المرجح $(\bar{w})$ يحسب بتطبيق المعادلة التالية :

$$\left(\overline{w}\right) = \frac{\sum xw}{\sum w} \qquad (٣-٨)$$

مزايا وعيوب الوسط الحسابي

يتميز الوسط الحسابي بالمزايا التالية :

- أنه سهل الحساب .
- يأخذ في الاعتبار كل القيم .
- أنه أكثر المقاييس استخداما وفهما .

ومن عيوبه .

- أنه يتأثر بالقيم الشاذة والمتطرفة .
- يصعب حسابه في حالة البيانات الوصفية .
- يصعب حسابه في حالة الجداول التكرارية المفتوحة .

2/2/3 الوسيط Median

هو أحد مقاييس النزعة المركزية، والذي يأخذ في الاعتبار رتب القيم ، ويعرف الوسيط بأنه القيمة التي يقل عنها نصف عدد القيم $(n/2)$ ، ويزيد عنها النصف الآخر $(n/2)$ ، أي أن 50% من القيم أقل منه، 50% من القيم أعلى منه. وفيما يلي كيفية حساب الوسيط في حالة البيانات غير مبوبة ، والبيانات المبوبة.

أولا: الوسيط للبيانات غير المبوبة

لبيان كيف يمكن حساب الوسيط للبيانات غير المبوبة ، نتبع الخطوات التالية:

- ترتب القيم تصاعديا .
- تحديد رتبة الوسيط، وهي : رتبة الوسيط = $\left(\frac{n+1}{2}\right)$
- إذا كان عدد القيم (n) فردي فإن الوسيط هو:

$$\text{الوسيط} = \text{القيمة رقم} \left(\frac{n+1}{2}\right) \qquad (٣-٩)$$

- إذا كان عدد القيم (n) زوجي، فإن الوسيط يقع بين القيمة رقم $(n/2)$، والقيمة رقم $((n/2)+1)$، ومن ثم يحسب الوسيط بتطبيق المعادلة التالي:

$$\text{الوسيط} = \frac{\text{القيمة رقم}\left(\frac{n}{2}\right) + \text{القيمة رقم}\left(\frac{n}{2}+1\right)}{2}$$ (٣-١٠)

مثـــال (3-3)

تم تقسيم قطعة أرض زراعية إلى 17 وحدة تجريبية متشابهة ، وتم زراعتها بمحصول القمح ، وتم استخدام نوعين من التسميد هما : النوع (a) وجرب على 7 وحدات تجريبية ، والنوع (b) وجرب على 10 وحدات تجريبية ، وبعد انتهاء الموسم الزراعي ، تم تسجيل إنتاجية الوحدة بالطن / هكتار ، وكانت على النحو التالي :

النوع (a)	1.2	2.75	3.25	2	3	2.3	1.5			
النوع (b)	4.5	1.8	3.5	3.75	2	2.5	1.5	4	2.5	3

والمطلوب حساب وسيط الإنتاج لكل نوع من السماد المستخدم، ثم قارن بينها.

الحـــل

أولا : حساب وسيط الإنتاج للنوع الأول (a)

- ترتيب القيم تصاعديا :

قيمة الوسيط

الإنتاج	1.2	1.5	2	2.3	2.75	3	3.25
الرتبة	1	2	2	4	5	6	7

رتبة الوسيط

- عدد القيم فردى $(n = 7)$
- إذا رتبة الوسيط هي: $\left((n+1)/2 = (7+1)/2 = 4\right)$.
- ويكون الوسيط هو القيمة رقم 4 ، أي أن وسيط الإنتاج للنوع a هو:

$Med_a = 2.3$ طن / هكتار

ثانيا : حساب وسيط الإنتاج للنوع الثاني (b) :

- ترتيب القيم تصاعديا .

$\text{قيمة الوسيط} = \frac{2.5+3}{2}$

2.75

الإنتاج	1.5	1.8	2	2.5	2.5	3	3.5	3.75	4	4.5
الرتبة	1	2	3	4	5	6	7	8	9	10

5.5

رتبة الوسيط

- عدد القيم زوجي $(n=10)$ إذا
- رتبة الوسيط هي : $\left((n+1)/2=(10+1)/2=5.5\right)$.
- الوسيط = الوسط الحسابي للقيمتين الواقعتين في المنتصف (رقم 5 ، 6) .

$$Med_b=\frac{2.5+3}{2}=2.75$$ طن / هكتار

وبمقارنة النوعين من السماد ، نجد أن وسيط إنتاجية النوع (a) أقل من وسيط إنتاجية النوع (b) ، أي أن : $Med_b > Med_a$.

ثانيا: الوسيط للبيانات المبوبة

لحساب الوسيط من بيانات مبوبة في جدول توزيع تكراري ، يتم إتباع الخطوات التالية .

- تكوين الجدول التكراري المتجمع الصاعد .
- تحديد رتبة الوسيط : $\left(\frac{n}{2}\right)=\left(\frac{\sum f}{2}\right)$
- تحديد فئة الوسيط كما في الشكل التالي :

تكرار متجمع صاعد سابق f_1		الحد الأدنى لفئة الوسيط (A)
رتبة الوسيط $(n/2)$	←	الوسيط Med
تكرار متجمع صاعد لاحق f_2		الحد الأعلى لفئة الوسيط

- ويحسب الوسيط ، بتطبيق المعادلة .

$$Med = A+\frac{\frac{n}{2}-f_1}{f_2-f_1}\times L \qquad (٣-١١)$$

حيث أن :

L هي طول فئة الوسيط، وتحسب بالمعادلة التالية:

طول الفئة = الحد الأعلى – الحد الأدنى

$$L = Upper - Lower$$

مثال (3-4)

فيما يلي توزيع 50 عجل متوسط الحجم ، حسب احتياجاته اليومية مـن الغـذاء الجـاف بالكيلوجرام

فئات الاحتياجات اليومية	1.5 -	4.5 -	7.5 -	10.5 -	13.5 – 16.5
عدد العجول f	4	12	19	10	5

والمطلوب : حساب الوسيط : أ - حسابيا ب- بيانيا

الحـــل

أولا : حساب الوسيط حسابيا

- رتبة الوسيط : $\frac{n}{2}=\frac{\sum f}{2}=\frac{50}{2}=25$
- الجدول التكراري المتجمع الصاعد :

تكرار متجمع صاعد	أقل من
0	1.5
4	4.5
f_1 16	A 7.5
25 ← رتبة الوسيط	Med (الوسيط)
f_2 35	10.5
45	13.5
50	16.5

- تحديد فئة الوسيط : وهى الفئة التي تشمل قيمة الوسيط ، وهي قيمة أقل منها $(n/2)$ من القيم ، ويمكن معرفتها بتحديد التكرارين المتجمعين الصاعدين الذين يقع بينهما رتبة الوسيط $(n/2)$ ، وفى الجدول أعلاه نجد أن رتبة الوسيط (25) تقع بين التكرارين المتجمعين (35 , 16) ، ويكون الحد الأدنى لفئة الوسيط هو المناظر للتكرار المتجمع الصاعد السابق 7.5 ، والحد الأعلى لفئة الوسيط هو المناظر للتكرار المتجمع الصاعد اللاحق 10.5 . أى أن فئة الوسيط هي : (7.5-10.5) .
- وبتطبيق معادلة الوسيط رقم (**3-11**) على هذا المثال نجد أن :

$$A=7.5\ ,\ f_1=16\ ,\ f_2=35\ ,\ L=10.5-7.5=3$$

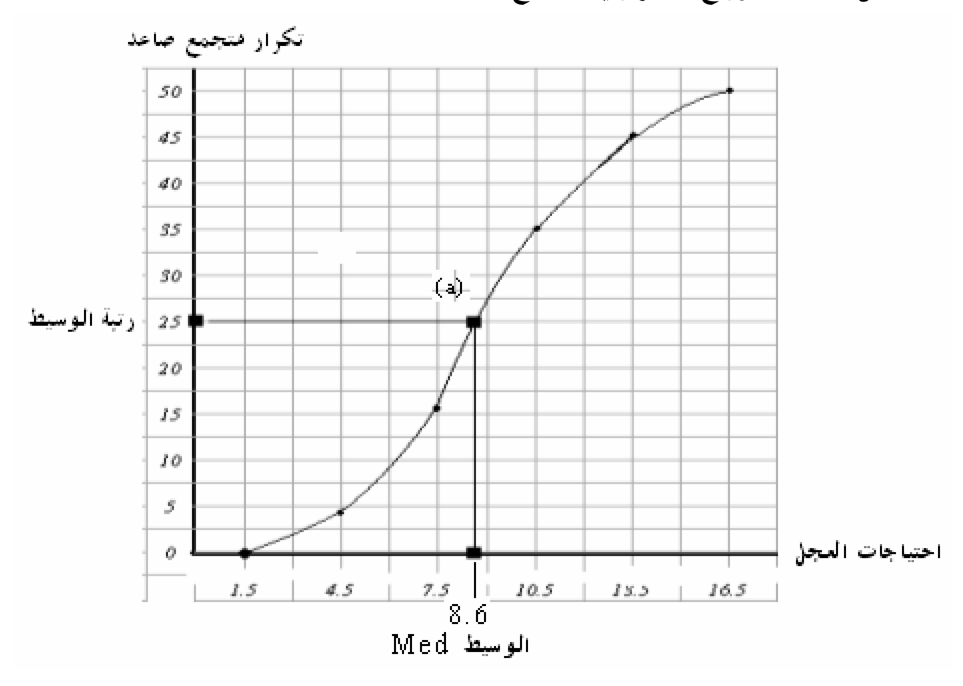
إذا الوسيط قيمته هي :

$$Med = A + \frac{\frac{n}{2} - f_1}{f_2 - f_1} \times L = 7.5 + \frac{25-16}{35-16} \times 3$$

$$= 7.5 + \frac{9}{19} \times 3 = 7.5 + \frac{27}{19} = 7.5 + 1.421 = 8.921 \ k.g$$

ثانيا :حساب الوسيط بيانيا

- تمثيل جدول التوزيع التكراري المتجمع الصاعد بيانيا .



- تحديد رتبة الوسيط (25) على المنحنى التكراري المتجمع الصاعد . ثم رسم خط مستقيم أفقي حتى يلقى المنحنى في النقطة (a) .
- إسقاط عمود رأسي من النقطة (a) على المحور الأفقي .
- نقطة تقاطع الخط الرأسي مع المحور الأفقي تعطى قيمة الوسيط .
- الوسيط كما هو مبين في الشكل Med = 8.6 .

مزايا وعيوب الوسيط

من مزايا الوسيط

1- لا يتأثر بالقيم الشاذة أو المتطرفة .

2- كما أنه سهل في الحساب .

3- مجموع قيم الانحرافات المطلقة عن الوسيط أقل من مجموع الانحرافات المطلقة عن أي قيم أخرى . أي أن : $\sum |x - Med| \le \sum |x - a| \ , \ a \ne Med$

ومن عيوب الوسيط

1- أنه لا يأخذ عند حسابه كل القيم في الاعتبار، فهو يعتمد على قيمة أو قيمتين فقط .

2- يصعب حسابه في حالة البيانات الوصفية المقاسة بمعيار اسمي nominal

3/2/3 المنوال Mode

يعرف المنوال بأنه القيمة الأكثر شيوعا أو تكرارا ، ويكثر استخدامه في حالة البيانات الوصفية ، لمعرفة النمط (المستوى) الشائع، ويمكن حسابة للبيانات المبوبة وغير المبوبة كما يلي:

أولا: حساب المنوال في حالة البيانات غير المبوبة

المنوال (Mod) = القيمة (المستوى) الأكثر تكرارا (٣-١٢)

ثانيا: حساب المنوال في حالة البيانات المبوبة (طريقة الفروق)

$$Mod = A + \frac{d_1}{d_1 + d_2} \times L \qquad (٣-١٣)$$

حيث أن :

A : الحد الأدنى لفئة المنوال (الفئة المناظرة لأكبر تكرار) .

d_1 : الفرق الأول = (تكرار فئة المنوال – تكرار سابق)

d_2 : الفرق الثاني = (تكرار فئة المنوال – تكرار لاحق)

L : طول فئة المنوال .

فئـــــة المنوال = الفئة المناظرة لأكبر تكرار

d_1 = (تكرار فئة المنوال – تكرار سابق)

d_2 = (تكرار فئة المنوال – تكرار سابق)

تكرار سابق

تكرار فئة المنوال

تكرار لاحق

مثـــال (3-5)

اختيرت عينات عشوائية من طلاب بعض أقسام كلية علوم الأغذية والزراعة ، وتم رصد درجات هؤلاء الطلاب في مقرر **122** إحصاء التطبيقي ، وكانت النتائج كالتالي:

قسم وقاية النباتات	80	77	75	77	77	77	65	70	58	67
قسم علوم الأغذية	88	68	60	75	93	65	77	85	95	90
قسم الاقتصاد	80	65	69	80	65	88	76	65	86	80
قسم الإنتاج الحيواني	85	73	69	85	73	69	69	73	72	85

والمطلوب حساب منوال الدرجات لكل قسم من الأقسام :

الحـــل

هذه البيانات غير مبوبة ، لذا فإن :

المنوال = القيمة الأكثر تكرارا

والجدول التالي يبين منوال الدرجة لكل قسم من الأقسام .

القيمة المنوالية	القيمة الأكثر تكرار	القسم
المنوال = 77 درجة	الدرجة 77 تكررت 4 مرات	قسم وقاية النباتات
لا يوجد منوال	جميع القيم ليس لها تكرار	قسم علوم الأغذية
يوجد منوالان هما : المنوال الأول = 65 المنوال الثاني = 80	الدرجة 65 تكررت 3 مرات الدرجة 80 تكررت 3 مرات	قسم الاقتصاد
يوجد ثلاث منوال هي : المنوال الأول = 69 المنوال الثاني = 73 المنوال الثالث = 85	الدرجة 69 تكررت 3 مرات الدرجة 73 تكررت 3 مرات الدرجة 85 تكررت 3 مرات	قسم الإنتاج الحيواني

مثال (3-6)

فيما يلي توزيع 30 أسرة حسب الإنفاق الاستهلاكي الشهري لها بالألف ريال .

فئات الإنفاق	2 -	5 -	8 -	11 -	14 - 17
عدد الأسر f	4	7	10	5	4

والمطلوب حساب منوال الإنفاق الشهري للأسرة، باستخدام طريقة الفروق .

الحل

لحساب المنوال لهذه البيانات يتم استخدام المعادلة رقم **(3-12)** ، ويتم إتباع الآتي :

- تحديد الفئة المنوالية

الفئة المنوالية هي الفئة المناظرة لأكبر تكرار : (8-11)

الفئات	التكرارات
2 -	4
5 -	7
8 -	10
11 -	5
14 - 17	4

فئة المنوال ← 8 - ، $A=8$

أكبر تكرار ← 10

$d_1 = 10-7 = 3$

$d_2 = 10-5 = 5$

- حساب الفروق d ، حيث أن :

$$d_1 = (10-7) = 3 \quad d_2 = (10-5) = 5$$

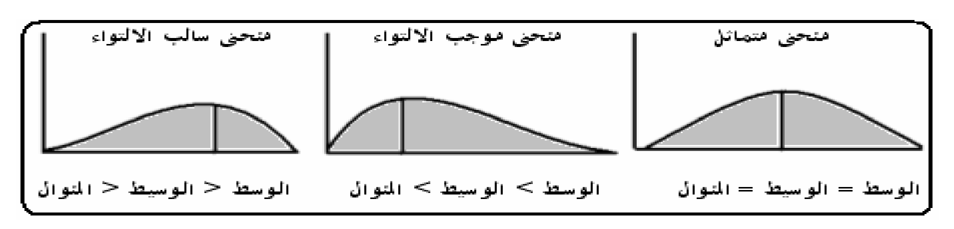
- تحديد الحد الأدنى للفئة المنوالية $(A=8)$ ، وكذلك طول الفئة $(L=3)$
- وبتطبيق المعادلة الخاصة بحساب المنوال فى حالة البيانات المبوبة . نجد أن :

$$Mod = A + \frac{d_1}{d_1 + d_2} \times L$$
$$= 8 + \frac{3}{3+5} \times 3 = 8 + 1.125 = 9.125$$

3/3 استخدام مقاييس التزعة المركزية في تحديد شكل توزيع البيانات

يمكن استخدام الوسط الحسابي والوسيط والمنوال في وصف المنحنى التكراري، والذي يعبر عن شكل توزيع البيانات ، كما يلي :

شكل **(3-1)**



- يكون المنحنى متماثل إذا كان :

الوسط = الوسيط = المنوال .

- يكون المنحنى موجب الالتواء (ملتوي جهة اليمين) إذا كان:

الوسط > الوسيط > المنوال

- يكون المنحنى سالب الالتواء (ملتوي جهة اليسار) إذا كان :

الوسط < الوسيط < المنوال

مثال عام **(3-7)**

قام مدير مراقبة الإنتاج بسحب عينة من 10 عبوات من المياه المعبأة للشرب ، ذات الحجم 5 لتر ، والمنتجة بواسطة إحدى شركات تعبئة المياه لفحص كمية الأملاح الذائبة، وكانت كالتالي :

115 123 119 123 124 119 123 121 123 121

والمطلوب : حساب الوسط الحسابي، والوسيط، والمنوال، ثم حدد شكل الالتواء لهذه البيانات .

الحل

حساب الوسط الحسابي :

$$\overline{X} = \frac{\sum x}{n} = \frac{1211}{10} = 121.1$$

- حساب الوسيط :

رتبة الوسيط : $(n+1)/2 = (10+1)/2 = 5.5$

ترتيب القيم تصاعديا

	قيمة الوسيط									
القيمة	115	119	119	121	121	123	123	123	123	124
الرتبة	1	2	3	4	5	6	7	8	9	10

قيمة الوسيط: 122 — رتبة الوسيط: 5.5

عدد القيم = 10 ، وهو عدد زوجي. الوسيط = الوسط الحسابي للقيمتين رقم (5 , 6)

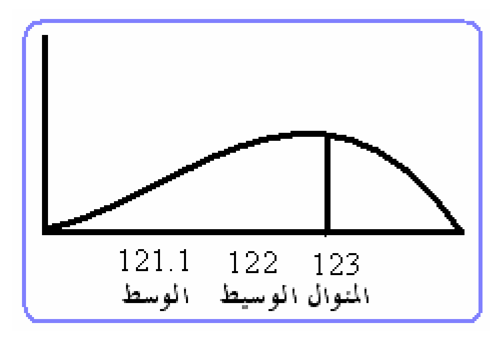
$$Med = \frac{121+123}{2} = \frac{244}{2} = 122$$

- حساب المنوال :

المنوال يساوى القيمة الأكثر تكرارا: القيمة 123 تكررت أكثر من غيرها ، إذا

$$Mod = 123$$

وبمقارنة الوسط والوسيط و المنوال نجد أن :



نجد أن : الوسط < الوسيط < المنوال ، إذا توزيع بيانات كمية الأملاح سالبة الالتواء.

مثال (3-8)

الجدول التكراري التالي يعرض توزيع 100 عامل في مزرعة حسب الأجر اليومي بالريال .

الأجر	50 -	70 -	90 -	110 -	130 -	150 -	170 - 190
عدد العمال	8	15	28	20	15	8	6

والمطلوب :

- **حساب الوسط والوسيط والمنوال .**
- **بيان شكل توزيع الأجور في هذه المزرعة .**

الحل

- **حساب الوسط والوسيط والمنوال .**

أولا : الوسط الحسابي $\overline{x}$

فئات الأجر	التكرارات (f)	مراكز الفئات (x)	f x
50 – 70	8	60	480
70 – 90	15	80	1200
90 – 110	28	100	2800
110 - 130	20	120	2400
130 - 150	15	140	2100
150 – 170	8	160	1280
170 - 190	6	180	1080
المجموع	100		11340

$$\overline{X} = \frac{\sum fx}{\sum f} = \frac{11340}{100} = 113.4 \;\; R.S$$

ثانيا : الوسيط Med

رتبة الوسيط : (n/2 =100/2 =50)

تكوين التوزيع التكراري المتجمع الصاعد .

أقل من	تكرار متجمع صاعد
أقل من 50	0
أقل من 70	8
أقل من 90	$23 \leftarrow f_1$
أقل من 110	$51 \leftarrow f_1$
أقل من 130	71
أقل من 150	86
أقل من 170	94
أقل من 190	100

رتبة الوسيط (50)

من الجدول أعلاه نجد أن :

$$\frac{n}{2}=50\ ,\ f_1=23\ ,\ f_2=51\ ,A=90\ ,L=110-90=20$$

إذا الوسيط قيمته هى :

$$Med=A+\frac{\frac{n}{2}-f_1}{f_2-f_1}\times L=90+\frac{50-23}{51-23}\times 20$$

$$=90+\frac{27}{28}\times 20=90+\frac{540}{28}=90+19.286=109.3\ \ R.S$$

ثالثا : المنوال Mod

الفئة المنوالية ، هى الفئة المناظرة لأكبر تكرار

أكبر تكرار = 28 ، وهو يناظر الفئة التقريبية (110 - 90) .

حساب الفروق : $d_1=28-15=13\quad ,\quad d_2=28-20=8$

الحد الأدنى للفئة : $A=90$ طول الفئة : $L=110-90=20$

إذا المنوال يحسب بتطبيق المعادلة التالية :

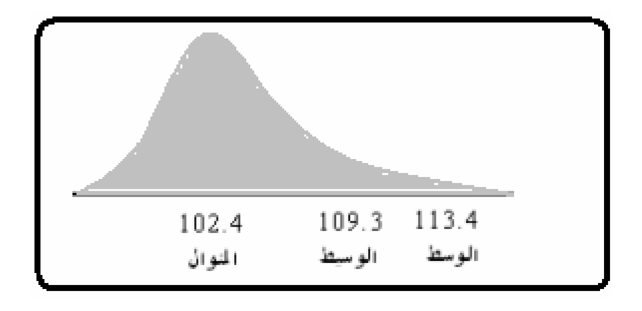
$$Mod=A+\frac{d_1}{d_1+d_2}\times L=90+\frac{13}{13+8}\times 20=90+\frac{260}{21}=1024\ \ R.S$$

- بيان شكل التوزيع .

من النتائج السابقة ، نجد أن :

الوسط الحسابي : $\overline{X}=113.4$ الوسيط: $Med=109.3$ المنوال: $Mod=1024$

أى أن : الوسط > الوسيط > المنوال إذا توزيع بيانات الأجور موجب الالتواء. كما هو مبين في الشكل التالي:



4/3 الرباعيات Quartiles

عند تقسيم القيم إلى أربع أجزاء متساوية، يوجد ثلاث إحصاءات ترتيبي تسمى بالرباعيات، وهي:

- الربيع الأول: وهو القيمة التي يقل عنها ربع عدد القيم، أي يقل عنها 25% من القيم، ويرمز له

بالرمز Q_1 .

- الربيع الثاني: وهو القيمة التي يقل عنها نصف عدد القيم، أي يقل عنها 50% من القيم، ويرمز له بالرمز Q_2 ، ومن ثم يعبر هذا الربيع عن الوسيط.
- الربيع الثالث: وهو القيمة التي يقل عنها ثلاث أرباع عدد القيم، أي يقل عنها 75% من القيم، ويرمز له بالرمز Q_3 .

والشكل (3-3) يبين أماكن الرباعيات الثلاث.

شكل (3-3)

الرباعيات

75% من القيم أقل من الربيع الثالث

50% من القيم أقل من البيع الثاني

25% من القيم أقل من الربع الأول

Q1 ربيع أول — Q2 ربيع ثاني — Q3 ربيع ثالث

ولحساب أي من الرباعيات الثلاث، يتم إتباع الآتي:

- بفرض أن عدد القيم عددها n، وأنها مرتبة كالتالي:

القيم مرتبة:	$X_{(1)}$ <	$X_{(2)}$ <	$X_{(3)}$	< $X_{(n)}$
: الرتبة	1	2	3	n

- تحديد رتبة الرباعي رقم i ، (Q_i) : $R=(n+1)\times\left(\frac{i}{4}\right)$
- إذا كانت R عددا صحيحا فإن قيمة الربيع هو: $Q_i = X_{(R)}$.
- إذا كانت R عدد كسري، فإن الرباعي (Q_i) يقع في المدى : $X_{(l)} < Q_i < X_{(u)}$ ، ومن ثم يحسب (Q_i) بالمعادلة التالية:

$$Q_i = x_{(l)} + (R-l)(x_{(u)} - x_{(l)}) \qquad (٣-١٤)$$

مثال (3-9)

فيما يلي كمية الإنتاج اليومي من الحليب باللتر للبقرة الواحدة لعينة حجمها 10 أبقار اختيرت مـــن مزرعة معينة:

25 23 29 32 34 29 20 18 27 30

احسب الرباعيات الثلاث لكمية الإنتاج، وما هو تعليقك؟

الحل:

لحساب الرباعيات الثلاث، يتم إتباع الآتي:

- ترتيب القيم تصاعديا:

قمة الربيع	22.25	28	30.5

القيم	18	20	23	25	27	29	29	30	32	34
الرتبة	1	2	3	4	5	6	7	8	9	10
رتبة الربيع		2.75			5.5			8.25		

- حساب الربيع الأول (Q_1):

رتبة الربيع الأول هي: $R = (n+1)\times\left(\frac{i}{4}\right) = (10+1)\times\left(\frac{1}{4}\right) = 2.75$

يقع الربيع الأول بين القيمتين: $(20 < Q_1 < 23)$، وبتطبيق المعادلة (**3-14**) نجد أن:

$$l = 2,\ \ R = 2.75\ \ , x_{(l)} = 20\ \ .x_{(u)} = 23$$

إذا :

$$Q_1 = x_{(l)} + (R-l)\times(x_{(u)} - x_{(l)}) = 20 + 0.75(23-20) = 22.25$$

- حساب الربيع الثاني (الوسيط) Q_2

رتبة الربيع الثاني هي: $R = (n+1)\times\left(\frac{i}{4}\right) = (10+1)\times\left(\frac{2}{4}\right) = 5.5$

يقع الربيع الثاني بين القيمتين: $(27 < Q_2 < 29)$، وبتطبيق المعادلة (**3-14**) نجد أن:

$$l = 5,\ \ R = 5.5\ \ , x_{(l)} = 27\ \ .x_{(u)} = 29$$

إذا :

$$Q_2 = x_{(l)} + (R-l)\times(x_{(u)} - x_{(l)}) = 27 + 0.5(29-27) = 28$$

- حساب الربيع الثالث Q_3

رتبة الربيع الثالث هي: $R = (n+1)\times\left(\frac{i}{4}\right) = (10+1)\times\left(\frac{3}{4}\right) = 8.25$

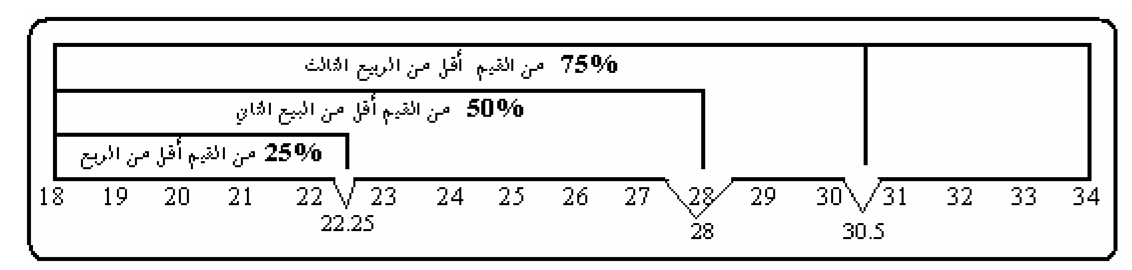
يقع الربيع الثالث بين القيمتين: $(30 < Q_3 < 32)$، وبتطبيق المعادلة (**3-14**) نجد أن:

$$l = 8,\ \ R = 8.25\ \ , x_{(l)} = 30\ \ .x_{(u)} = 32$$

إذا :

$$Q_3 = x_{(l)} + (R-l)\times(x_{(u)} - x_{(l)}) = 30 + 0.25(32-30) = 30.5$$

من النتائج السابقة نجد أن:



- 25% من الأبقار يقل إنتاجه عن 22.25 لتر يوميا.
- 50% من الأبقار يقل إنتاجه عن 28 لتر يوميا.
- 75% من الأبقار يقل إنتاجه عن 30.5 لتر يوميا.

تمارين

أولا : استخدم البيانات التالية ، ثم أجب عما هو مطلوب باختيار الإجابة الصحيحة من بين الإجابات الأربعة : فيما يلى الطاقة التصديرية من المياه بالألف كيلومتر مكعب يوميا (x) ، لـــعدد 10 محطات تحلية .

x: 342 216 105 291 107 216 210 165 90 216

1- هذه البيانات من النوع :

(a) الكمى المنفصل (b) الكمى المتصل (c) الوصفى (d) الوصفى الترتيبى

2- $\sum x$ قيمتها:

(a) 1000 (b) 1958 (c) 195.8 (d) 216

3- قيمة الطاقة التصديرية التى أقل منها 50% من القيم تسمى :

(a) الوسيط (b) الوسط (c) التباين (d) المدى

4- القيمة الأكثر تكرارا تسمى :

(a) الوسيط (b) الوسط (c) المنوال (d) الانحراف

5- الوسط الحسابى للطاقة التصديرية قيمته :

(a) 216 (b) 1958 (c) 195.8 (d) 213

6- المنوال قيمته

(a) 216 (b) 1958 (c) 195.8 (d) 347

7- الوسيط قيمته

(a) 213 (b) 1958 (c) 195.8 (d) 216

8- تعتبر بيانات الطاقة التصديرية أعلاه لها توزيع

(a) متماثل (b) سالب الالتواء (c) موجب الالتواء (d) غير معروف .

9- إذا تم إدخال تعديل على هذه المحطات لزيادة الطاقة التصديرية لكل محطة 50 ألف كيلو متر مكعب ، يكون الوسط الحسابى للطاقة التصديرية بعد التطوير هو .

(a) 216 (b) 1958 (c) 195.8 (d) 245.8

10- إذا كانت $y = 0.5x$ فإن الوسط الحسابى للقيم التى يأخذها المتغير الجديد y هو :

(a) 216 (b) 97.9 (c) 195.8 (d) 245.8

ثانيا : فيما يلى التوزيع التكرارى لــعدد 50 مزرعة حسب المساحة المترزعة بمحصول الطماطم بالألف دونم .

المساحة بالألف دونم	4.5 –	7.5 –	10.5 -	13.5 -	16.5-	19.5 – 22.5
عدد المزارع	3	8	12	15	10	2

استخدم بيانات الجدول أعلاه للإجابة على الأسئلة من (11- 20)

11- طول الفئة قيمته

(a) 1 (b) 2 (c) 3 (d) 5

12- الحد الأدنى للفئة الرابعة هو

(a) 14.5 (b) 16 (c) 15 (d) 13.5

13- مركز الفئة الثانية قيمته

(a) 9 (b) 8 (c) 10 (d) 3

14- مجموع التكرار النسبى للفئات يساوى :

(a) 0.30 (b) 0.20 (c) 1 (d) 1.50

15- إذا كانت x هى مركز الفئة ، f هو تكرار الفئة فإن $\sum fx$ قيمته تساوى

(a) 225 (b) 225 (c) 50 (d) 681

16- الوسط الحسابى قيمته تساوى

(a) 8.33 (b) 13.5 (c) 13.62 (d) 681

17- الفئة التى يقع فيها قيمة الوسيط هى :

(a) 13.5 – 16.5 (b) 16.5- 19.5 (c) 14 – 17 (d) 10.5 – 13.5

18- رتبة الوسيط هى :

(a) 50 (b) 10 (c) 25 (d) 1

19- الوسيط قيمته تساوى .

(a) 13.9 (b) 13.5 (c) 15 (d) 12.5

20- المنوال قيمته تساوى :

(a) 14 (b) 15 (c) 13.5 (d) 14.625

21- من الإجابة **16** ، **19** ، **20** يكون شكل التوزيع .

(a) ملتوى جهة (b) متماثل (c) سالب (d) غير محدد

اليمين الإلتواء

ثالثا : قم بتسجيل البيانات التالية :

الإسم : الرقم الجامعي:

قم بتظليل الاختيار الصحيح من (1 – 21) ، ولا ينظر للإجابة التى بها مربعين مظللين :

رقم السؤال	(a)	(b)	(c)	(d)
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				

الفصـــــــــــــل الرابــــــع
مقاييس التشتت

1/4 مقدمة

عند مقارنة مجموعتين من البيانات ، يمكن استخدام شكل التوزيع التكراري، أوالمنحنى التكراري ، وكذلك بعض مقاييس النزعة المركزية ، مثل الوسط الحسابي والوسيط ، والمنوال ، والإحصاءات الترتيبية ، ولكن استخدام هذه الطرق وحدها لا يكفي عند المقارنة ، فقد يكون مقياس النزعة المركزية للمجموعتين متساوي ، وربما يوجد اختلاف كبير بين المجموعتين من حيث مدى تقارب وتباعد البيانات من بعضها البعض ، أو مدى تباعد أو تقارب القيم عن مقياس النزعة المركزية .

ومثال على ذلك ، إذا كان لدينا مجموعتين من الطلاب ، وكان درجات المجموعتين كالتالي :

المجموعة الأولى	63	70	78	81	85	67	88
المجموعة الثانية	73	78	77	78	75	74	77

لو قمنا بحساب الوسط الحسابي لكل مجموعة ، نجد أن الوسط الحسابي لكل منهما يساوي 76 درجة ، ومع ذلك درجات المجموعة الثانية أكثر تجانسا من درجات المجموعة الأولى . من أجل ذلك لجأ الإحصائيون إلى استخدام مقاييس أخرى لقياس مدى تجانس البيانات، أو مدى انتشار البيانات حول مقياس النزعة المركزية، ويمكن استخدامها في المقارنة بين مجموعتين أو أكثر من البيانات، ومن هذه المقاييس ، مقاييس التشتت ، والالتواء ، والتفرطح ، وسوف نركز في هذا الفصل على هذه المقاييس .

2/4 مقاييس التشتت Dispersion Measurements

من هذه المقاييس: المدى، والانحراف الربيعي، والانحراف المتوسط، والتباين، والانحراف المعياري .

1/2/4 المدى Rang

هو أبسط مقاييس التشتت ، ويحسب المدى في حالة البيانات غير المبوبة بتطبيق المعادلة التالية .

المدى في حالة البيانات غير المبوبة = أكبر قراءة − أقل قراءة

$$Rang = Max - Min \quad (\text{٤}-\text{١})$$

وأما المدى في حالة البيانات المبوبة له أكثر من صيغة، ومنها المعادلة التالية:

(٤-٢) المدى في حالة البيانات المبوبة = مركز الفئة الأخيرة – مركز الفئة الأولى

مثـــــال (4-1)

تم زراعة 9 وحدات تجريبية بمحصول القمح ، وتم تسميدها بنوع معين من الأسمدة الفسفورية ، وفيما يلي بيانات كمية الإنتاج من القمح بالطن/ هكتار .

4.8 6.21 5.4 5.18 5.29 5.18 5.08 4.63 5.03

والمطلوب حساب المدى .

الحـــل

المدى = أكبر قراءة – أقل قراءة

أكبر قراءة = 6.21 أقل قراءة = 4.63

إذا المدى هو :

Rang=Max-Min=6.21-4.63 =1.58

المدى يساوي 1.58 طن / هكتار.

مثـــال (4-2)

الجدول التكراري التالي يبين توزيع 60 مزرعة حسب المساحة المتزرعة بالذرة بالألف دونم .

المساحة	15-20	20-25	25-30	30-35	35-40	40-45
عدد المزارع	3	9	15	18	12	3

والمطلوب حساب المدى للمساحة المتزرعة بالذرة .

الحـــل

المدى = مركز الفئة الأخيرة – مركز الفئة الأولى

مركز الفئة الأخيرة: (40+45)/2=85/2=42.5 مركز الفئة الأولى: (15+20)/2=35/2=17.5

إذا $$Rang = 42.5 - 17.5 = 25$$

أي أن المدى قيمته تساوي 25 دونم

مزايا وعيوب المدى

من مزايا المدى

1- أنه بسيط وسهل الحساب
2- يكثر استخدامه عند الإعلان عن حالات الطقس، و المناخ الجوي، مثل درجات الحرارة، والرطوبة، والضغط الجوي.
3- يستخدم في مراقبة الجودة .

2- ومن عيوبه

- أنه يعتمد على قيمتين فقط ، ولا يأخذ جميع القيم في الحسبان .
- يتأثر بالقيم الشاذة .

2/2/4 الانحراف الربيعي Quartile Deviation (Q)

يعتمد المدى على قيمتين متطرفتين ، هما أصغر قراءة ، وأكبر قراءة ، فإذا كان هناك قيم شاذة، ترتب على استخدامه كمقياس للتشتت نتائج غير دقيقة، من أجل ذلك لجأ الإحصائيون، إلى استخدام مقياس للتشتت يعتمد على نصف عدد القيم الوسطى، ويهمل نصف عدد القيم المتطرفة، ولذا لا يتأثر هذا المقياس بوجود قيم شاذة، ويسمى هذا المقياس بالانحراف الربيعي (Q)، ويحسب الانحراف الربيعي بتطبيق المعادلة التالية .

$$Q = \frac{Q_3 - Q_1}{2} \qquad (٤-٣)$$

حيث أن Q_1 هو الربيع الأول ، Q_3 هو الربيع الثالث ، وقد بينا في الفصل الثالث كيف يمكن حساب هذان الرباعيان ، ومن المعادلة أعلاه ، يعرف الانحراف الربيعي بنصف المدى الربيعي ، أي أن :

الانحراف الربيعي = نصف المدى الربيعي

مثـــــال(4-3)

استخدم بيانات مثال (4-1) ، ثم احسب الانحراف الربيعي لكمية الإنتاج من القمح .

الحـــل

- ترتيب القيم تصاعديا

الإنتاج	4.63	4.8	5.03	5.08	5.18	5.18	5.29	5.4	6.21
الرتبة	1	2	3	4	5	6	7	8	9

- حساب الربيع الأول Q_1

رتبة الربيع الأول: $(n+1)\left(\frac{1}{4}\right) = (9+1)(0.25) = 2.5$.

$$x_{(l)} = x_{(2)} = 4.8 \ , \ x_{(u)} = x_{(3)} = 5.03 \ , \ R = 2.5 \ , \ l = 2 \ , R - l = 0.5$$

إذا

$$Q_1 = x_{(l)} + (r-l)(x_{(u)} - x_{(l)})$$
$$= 4.8 + 0.5(5.03 - 4.8) = 4.915$$

- حساب الرباعي الثالث (Q_3)

موقع الرباعي الثالث: $(n+1)\left(\frac{3}{4}\right)=(9+1)(0.75)=7.5$

$$x_{(l)}=x_{(7)}=5.29\ ,\ x_{(u)}=x_{(8)}=5.4\ ,\ R=7.5\ ,\ l=7\ ,R-l=0.5$$

إذا

$$Q_3=x_{(l)}+(R-l)(x_{(u)}-x_{(l)})$$
$$=5.29+0.5(5.4-5.29)=5.345$$

- حساب الانحراف الربيعي

$$Q=\frac{Q_3-Q_1}{2}=\frac{5.345-4.915}{2}=0.215$$

إذا الانحراف الربيعي قيمته تساوي 0.215 طن/ هكتار .

مثـــــال(4-4)

استخدم بيانات مثال رقم (**4-2**) في حساب نصف المدى الربيعي .

الحـــــــل:

عند حساب الربيع الأول أو الثالث يتبع نفس الأسلوب المستخدم في حساب الوسيط.

- تكوين الجدول التكراري المتجمع الصاعد
- حساب الرباعي الأول (Q_1)

رتبة الربيعي الأول : $n(1/4)=60(0.25)=15$

$$f=15\ ,\ f_1=12\ ,\ f_2=27\ ,\ A=25\ ,\ L=5$$

إذا

$$Q_1=A+\frac{f-f_1}{f_2-f_1}L$$
$$=25+\frac{15-12}{27-12}(5)=25+\frac{3}{15}(5)=26$$

حدود المساحة	عدد المزارع
15-	3
20-	9
25-	15
30-	18
35-	12
40-45	3
sum	60

أقل من	تكرار متجمع
15	0
20	3
A 25	f_1 12
30	f_2 27 (15)
A 35	45 (45)
40	57
45	60

- حساب الرباعي الثالث (Q_3)

موقع الرباعي الثالث : $n(3/4)=60(0.75)=45$

$$f=45\ ,\ f_1=45\ ,\ f_2=57\ ,\ A=35\ ,\ L=5$$

إذا

$$Q_3 = A + \frac{f - f_1}{f_2 - f_1}L$$
$$= 35 + \frac{45-45}{57-45}(5) = 35 + \frac{(0)}{15}(5) = 35$$

- نصف المدى الربيعي .

$$Q = \frac{Q_3 - Q_1}{2} = \frac{35-26}{2} = 4.5$$

إذا الانحراف الربيعي للمساحة 4.5 ألف دونم.

مزايا وعيوب الانحراف الربيعي

من مزايا الانحراف الربيعي، يفضل استخدامه كمقياس للتشتت في حالة وجود قيم شاذة ، كما أنه بسيط وسهل في الحساب . ومن عيوبه ، أنه لا يأخذ كل القيم في الاعتبار .

3/2/4 الانحراف المتوسط Mean Deviation (MD)

هو أحد مقاييس التشتت، ويعبر عنه بمتوسط الانحرافات المطلقة للقيم عن وسطها الحسابي ، فإذا كانت $x_1, x_2, ..., x_n$ هي القراءات التي تم أخذها عن ظاهرة معينة ، وكان ($\bar{x} = \sum x/n$) عبارة عن الوسط الحسابي لهذه القراءات، فإن الانحراف المتوسط (MD) يحسب بتطبيق المعادلة التالية:

$$MD = \frac{\sum |x - \bar{x}|}{n} \qquad (٤-٤)$$

وهذه الصيغة تستخدم في حالة البيانات غير المبوبة .

مثـــال (4-5)

إذا كانت الطاقة التصديرية لخمس محطات لتحلية المياه بالمليون متر مكعب كما يلي:

4 5 2 10 7

أوجد قيمة الانحراف المتوسط للطاقة التصديرية

الحـــل

لحساب قيمة الانحراف المتوسط يتم استخدام المعادلة (4-4)

- الوسط الحسابي :

$$\bar{x} = \frac{\sum x}{n} = \frac{28}{5} = 5.6$$

ويتم تكوين الجدول التالي :

الطاقة التصديرية x	الانحرافات $(x-\bar{x})=(x-5.6)$	الانحرافات المطلقة $\lvert x-5.6\rvert$
4	4 - 5.6 = -1.6	1.6
5	5 - 5.6 = -0.6	0.6
2	2 - 5.6 = -3.6	3.6
10	10 - 5.6 = 4.4	4.4
7	7 - 5.6 = 1.4	1.4
Sum	0	11.6

- إذا الانحراف المتوسط قيمته هي :

$$MD=\frac{\sum\lvert x-\bar{x}\rvert}{n}=\frac{11.6}{5}=2.32$$ (مليون متر مكعب)

وفي حالة البيانات المبوبة، يحسب الانحراف المتوسط باستخدام المعادلة التالية .

$$MD=\frac{\sum\lvert x-\bar{x}\rvert\, f}{n} \qquad (٤-٥)$$

حيث أن f هو تكرار الفئة ، x هو مركز الفئة ، $\bar{x}$ هو الوسط الحسابي.

مثـــال(4-6)

يبين الجدول التكراري التالي توزيع 40 أسرة حسب الإنفاق الشهري بالألف ريال.

الإنفاق	2 - 5	5 - 8	8 - 11	11 – 14	14 – 17
عدد الأسرة	1	8	13	10	8

أوجد الانحراف المتوسط .

الحــــــــــل

لحساب الانحراف المتوسط ، يتم تطبيق المعادلة (4-5)، ويتبع الآتي

- تكوين جدول لحساب مكونات المعادلة:

حدود الإنفاق	عدد الأسر f	مركز الفئة x	$x\ f$	الوسط الحسابي $\bar{x}$	$\lvert x-\bar{x}\rvert$	$\lvert x-\bar{x}\rvert\ f$
2-5	1	3.5	3.5	$\bar{x}=\frac{\sum x}{n}=\frac{428}{40}=10.7$	7.2	7.2
5-8	8	6.5	52		4.2	33.6
8-11	13	9.5	123.5		1.2	15.6
11-14	10	12.5	125		1.8	18
14-17	8	15.5	124		4.8	38.4
sum	40		428			112.8

إذا الانحراف المتوسط هو :

$$MD = \frac{\sum |x-\bar{x}|\ f}{n} = \frac{112.8}{40} = 2.82$$

الانحراف المتوسط للإنفاق الشهري هو 2.82 ألف ريال .

مزايا وعيوب الانحراف المتوسط

من مزايا الانحراف المتوسط أنه يأخذ كل القيم في الاعتبار، ولكن يعاب عليه ما يلي:

- يتأثر بالقيم الشاذة .
- يصعب التعامل معه رياضيا.

4/2/4 التباين Variance

هو أحد مقاييس التشتت ، وأكثرها استخداما في النواحي التطبيقية ، ويعبر عـــن متوســـط مربعات انحرافات القيم عن وسطها الحسابي.

أولا: التباين في المجتمع (s^2)

إذا توافر لدينا قراءات عن كل مفردات المجتمع ، ولتكن: $x_1, x_2, \ldots, x_N$ ، فإن التبـــاين في المجتمع ، ويرمز له بالرمز s^2 (سيجما) يحسب باستخدام المعادلة التالية :

$$\sigma^2 = \frac{\sum (x-u)^2}{N} \qquad (٤-٦)$$

حيث أن m هو الوسط الحسابي في المجتمع ، أى أن : $m = \sum x/N$.

مثـــال(4-7)

مصنع لتعبئة المواد الغذائية ، يعمل به 15 عامل ، وكانت عدد سنوات الخبرة لهؤلاء العمال كما يلي :

5 13 7 14 12 9 6 8 10 13 14 6 11 12 10

بفرض أن هذه البيانات تم جمعها عن كل مفردات المجتمع ، فأوجد التباين لعدد سنوات الخبرة .

الحـــل

لحساب تباين سنوات الخبرة في المجتمع ، يتم استخدام المعادلة (**4-6**).

- الوسط الحسابي في المجتمع m

$$m = \frac{1}{N}\sum x$$

$$= \frac{1}{15}(5+13++7+\ldots+12+10) = \frac{1}{15}(150) = 10$$

- حساب مربعات الانحرافات $\sum(x-m)^2$

بما أن: $\sum(x-m)^2 = 130$

إذا تباين سنوات الخبرة للعمال في المصنع هو :

$$s^2 = \frac{\sum(x-u)^2}{N} = \frac{130}{15} = 8.67$$

سنوات الخبرة x	$(x-m)$	$(x-m)^2$
5	5-10 = -5	25
13	3	9
7	-3	9
14	4	16
12	2	4
9	-1	1
6	-4	16
8	-2	4
10	0	0
13	3	9
14	4	16
6	-4	16
11	1	1
12	2	4
10	0	0
150	0	130

ويمكن تبسيط المعادلة (**4-6**) في صورة أخرى كما يلي :

يمكن فك المجموع $\sum(x-m)^2$ كالتالي :

$$\begin{aligned}\sum(x-m)^2 &= \sum\left(x^2 - 2xm + m^2\right)\\ &= \sum x^2 - 2m\sum x + \sum m^2\\ &= \sum x^2 - 2Nm^2 + Nm^2\\ &= \sum x^2 - Nm^2\end{aligned}$$

ومن ثم يكتب تباين المجتمع على الصورة التالية :

$$s^2 = \frac{\sum x^2 - Nm^2}{N} = \frac{1}{N}\sum x^2 - m^2$$

إذا التباين في المجتمع يمكن صياغته كالتالي .

$$\sigma^2 = \frac{1}{N}\sum x^2 - \mu^2 \qquad (٤-٧)$$

وبالتطبيق على المثال (**4-7**) ، نجد أن أننا نحتاج إلى المجموعين : $\sum x^2$, $\sum x$ ، ويتم عمل الآتي :

سنوات الخبرة x	x^2
5	25
13	169
7	49
14	196
12	144
9	81
6	36
8	64
10	100
13	169
14	196
6	36
11	121
12	144
10	100
150	1630

$$\sum x = 150 \;,\; \sum x^2 = 1630$$

$$m = \frac{1}{N}\sum x = \frac{1}{15}(150) = 10$$

إذا التباين هو

$$s^2 = \frac{1}{N}\sum x^2 - m^2$$

$$= \frac{1}{15}1630 - 10^2 = 108.67 - 100 = 8.67$$

وهي نفس النتيجة التي تم الحصول عليها باستخدام الصيغة (**4-6**) .

ثانيا: التباين في العينة (s^2)

في كثير من الحالات يكون تباين المجتمع s^2 غير معلوم، وعندئذ يتم سحب عينة من هذا المجتمع ، ويحسب التباين من بيانات العينة كتقدير لتباين المجتمع ، فإذا كانت قراءات عينة عشوائية حجمها n هي ، $x_1, x_2, \dots, x_n$ ، فإن تباين العينة ويرمز له بالرمز s^2 هو:

$$s^2 = \frac{\sum (x - \bar{x})^2}{n-1} \qquad (٤-٨)$$

حيث أن $\bar{x}$ هو الوسط الحسابي لقراءات العينة ، أي أن : $\bar{x} = \sum x/n$ ، وتباين العينة المبين بالمعادلة (**4-8**) هو التقدير غير المتحيز لتباين المجتمع .

مثـــال(4 - 8)

في المثال (4 - 7) السابق ، إذا تم سحب عينة من عمال المصنع حجمها 5 عمال ، وسجل عدد سنوات الخبرة ، وكانت كالتالي .

8 13 10 5 9

احسب تباين سنوات الخبرة في العينة .

الحــــــل

لحساب التباين في العينة يتم تطبيق المعادلة (4 - 8)، ويتبع الآتي :

- الوسط الحسابي في العينة :

$$\bar{x} = \frac{1}{n}\sum x = \frac{1}{5}(8+13+10+5+9) = \frac{1}{5}(45) = 9$$

- حساب مربعات الانحرافات $\sum (x - \bar{x})^2$

سنوات الخبرة x	8	13	10	5	9	45
$(x - \bar{x})$	-1	4	1	-4	0	0
$(x - \bar{x})^2$	1	16	1	16	0	34

أي أن : $\sum (x - \bar{x})^2 = 34$ ،

- إذا تباين سنوات الخبرة في العينة قيمته هي :

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1} = \frac{34}{(5 - 1)} = \frac{34}{4} = 8.5$$

- في هذه الحالة يمكن القول بأن تباين العينة 8.5، وهو في نفس الوقت تقدير غير متحيز لتباين المجتمع .

تبسيط العمليات الحسابية

يمكن تبسيط الصيغة الرياضية لتباين العينة الموضحة بالمعادلة (4 - 8) إلى صيغة سهلة يمكـــن التعامل معها، وخاصة إذا كانت البيانات تحتوي على قيم كسرية، ولاستنتاج هذه الصيغة يتم إتبـــاع الآتي.

يمكن فك المجموع $\sum (x - \bar{x})^2$ كالتالي:

$$\sum(x-\overline{x})^2=\sum\left(x^2-2x\overline{x}+\overline{x}^2\right)$$
$$=\sum x^2-2\overline{x}\sum x+\sum\overline{x}^2$$
$$=\sum x^2-2n\overline{x}^2+n\overline{x}^2$$
$$=\sum x^2-n\overline{x}^2$$

ويكتب تباين العينة على الصورة التالية :

$$s^2=\frac{1}{n-1}\left(\sum x^2-n\overline{x}^2\right)$$

إذا التباين في العينة يمكن صياغته كالتالي .

$$s^2=\frac{1}{n-1}\left(\sum x^2-n\overline{x}^2\right) \qquad (٤-٩)$$

كما يمكن إثبات أن المعادلة (**4-9**) تأخذ الشكل التالي:

$$s^2=\frac{1}{n-1}\left(\sum x^2-\frac{(\sum x)^2}{n}\right) \qquad (٤-١٠)$$

وبالتطبيق على بيانات المثال السابق ، نجد أن :

سنوات الخبرة X	8	13	10	5	9	45
x^2	64	169	100	25	81	439

- تباين العينة باستخدام المعادلة (**4-9**) هو :

$$s^2=\frac{1}{n-1}\left(\sum x^2-n\overline{x}^2\right)$$
$$=\frac{1}{5-1}\left(439-5(9)^2\right)=\frac{1}{4}(34)=8.5$$

- وباستخدام المعادلة (**4-10**) نجد أن:

$$s^2=\frac{1}{n-1}\left(\sum x^2-\frac{(\sum x)^2}{n}\right)$$
$$=\frac{1}{5-1}\left(439-\frac{(45)^2}{5}\right)=\frac{1}{4}(439-405)=\frac{1}{4}(34)=8.5$$

الانحراف المعياري Standard Deviation

عند استخدام التباين كمقياس من مقاييس التشتت، نجد أنه يعتمد علـــي مجمـــوع مربعـــات الانحرافات، ومن ثم لا يتمشى هذا المقياس مع وحدات قياس المتغير محل الدراسة ، ففي المثال السابق ، نجد أن تباين سنوات الخبرة في العينة 8.5، فليس من المنطق عند تفسير هذه النتيجة أن نقول ، " تباين سنوات الخبرة هو 8.5 سنة تربيع "، لأن وحدات قياس المتغير هو عدد السنوات، من أجل ذلك لجـــأ الإحصائيين إلى مقياس منطقي يأخذ في الاعتبار الجذر التربيعي للتباين ، لكي يناسب وحدات قيـــاس المتغير، وهذا المقياس هو الانحراف المعياري.

إذا الانحراف المعياري ، هو الجذر التربيعي الموجب للتباين ، أي أن:

$$\text{الانحراف المعياري} = \sqrt{\text{التباين}} \qquad (٤-١١)$$

ومثال على ذلك :

- في مثال (4-7) نجد أن الانحراف المعياري لسنوات الخبرة لعمال المصنع (المجتمع) ، ويرمز له بالرمز (s) هو :

$$s = \sqrt{\frac{1}{N}\sum x^2 - m^2}$$

$$= \sqrt{\frac{1}{15}1630 - 10^2} = \sqrt{8.67} = 2.94$$

في هذه الحالة ، يكون الانحراف المعياري لسنوات الخبرة في المجتمع هو 2.94 سنة .

- في مثال (4-8) نجد أن الانحراف المعياري لسنوات الخبرة لعمال العينة ، ويرمز له بالرمز S ، هو :

$$s = \sqrt{\frac{1}{n-1}\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)}$$

$$= \sqrt{\frac{1}{5-1}\left(439 - \frac{(45)^2}{5}\right)} = \sqrt{\frac{1}{4}(439-405)} = \sqrt{\frac{1}{4}(34)} = 2.92$$

أي أن الانحراف المعياري لسنوات الخبرة في العينة هو 2.92 سنة .

الانحراف المعياري في حالة البيانات المبوبة

إذا كانت بيانات الظاهرة ، مبوبة في جدول توزيع تكراري ، فإن الانحراف المعياري يحسب بتطبيق المعادلة التالية .

$$s = \sqrt{\frac{\sum (x-\bar{x})^2 f}{n-1}}$$

or

$$s = \sqrt{\frac{\sum x^2 f - \frac{\left(\sum xf\right)^2}{n}}{n-1}} \qquad (١٢-٤)$$

حيث أن f هو تكرار الفئة ، x هو مركز الفئة ، $\bar{x}$ هو الوسط الحسابي $\left(\sum xf/n\right)$ ، n هي مجموع التكرارات $\left(n=\sum f\right)$ ، والمقدار الذي تحت الجذر يعبر عن التباين $\left(s^2\right)$.

مثـــال(4-9)

في بيانات مثال (4-6) ، احسب الانحراف المعياري للإنفاق الشهري للأسرة ، ثم قارن بـــين الانحراف المتوسط ، والانحراف المعياري للإنفاق الشهري للأسرة .

الحـــــــــــل

لحساب الانحراف المعياري للإنفاق الشهري ، تستخدم المعادلة رقـــم (4-12)، وسوف نطبق الصيغة الثانية ، ولذا نكون جدول لحساب المجموعين : $\sum x^2 f$, $\sum xf$.

الإنفاق	عدد الأسر f	مركز الفئة x	xf	$x^2 f$
2-5	1	3.5	3.5	12.25
5-8	8	6.5	52	338
8-11	13	9.5	123.5	1173.25
11-14	10	12.5	125	1562.5
14-17	8	15.5	124	1922
sum	40		428	5008

$$n=\sum f = 40$$
$$\sum xf = 428$$
$$\sum x^2 f = 5008$$

وبتطبيق المعادلة ، نجد أن الانحراف المعياري قيمته هي :

$$s=\sqrt{\frac{\sum x^2 f-\frac{\left(\sum xf\right)^2}{n}}{n-1}}$$

$$=\sqrt{\frac{5008-\frac{(428)^2}{40}}{40-1}}=\sqrt{\frac{5008-4579.6}{39}}$$

$$=\sqrt{10.984615}=3.314$$

أي أن الانحراف المعياري للإنفاق الشهري 3.314 ألف ريال ، ووفقا لهذا المقيـــاس ، فـــإن تشتت بيانات الإنفاق أكبر من تشتت بيانات الإنفاق وفقا لمقياس الانحراف المتوسط (2.88) .

خصائص الانحراف المعياري

من خصائص الانحراف المعياري ، ما يلي :

- أولا : الانحراف المعياري للمقدار الثابت يساوي صفرا ، أي أنه إذا كان لدينا القراءات التالية: $x: a, \ a, a, \dots, a$ حيث أن a مقدار ثابت فإن : $S_x=0$ ، حيث أن S_x تعبر عن الانحـــراف المعياري لقيم x .

- ثانيا : إذا أضيف مقدار ثابت إلى كل قيمة من قيم المفردات ، فإن الانحراف المعياري للقيم الجديدة (القيم بعد الإضافة) تساوي الانحراف المعياري للقيم الأصلية (القيم بعد الإضافة) ، فإذا كانـــت القيم الأصلية هي $x_1, x_2, \dots, x_n$ ، وتم إضافة مقدار ثابت a إلى كل قيمة من قيم x ، فإن الانحراف المعياري للقيم الجديـــدة : $x_1+a, x_2+a, \dots, x_n+a$: $(y=x+a)$ هـــي : $S_y=S_x$:

مثـــال(4-10)

إذا كان من المعلوم أن تطبيق برنامج غذائي معين للتسمين لفترة زمنية محددة سوف يزيد من وزن الدجاجة 0.5 كيلوجرام، سحبت عينة عشوائيا من مزرعة دجاج حجمها 5 دجاجات، وكانت أوزانها كالتالي: **2.5 , 1.25 , 2 , 1.75 , 1** .

1- احسب الانحراف المعياري لوزن الدجاجة.

2- إذا طبق البرنامج الغذائي المشار إليه، ما هو الانحراف المعياري لوزن الدجاجة في هـــذه العينة؟

الحـــــــــــل

1- حساب الانحراف المعياري للوزن قبل تطبيق البرنامج .

x	x^2
1	1

$n=5$

1.75	3.0625
2	4
1.25	1.5625
2.5	6.25
8.5	15.875

إذا الانحراف المعياري للوزن قبل البرنامج في العينة هو:

$$s_x = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}}$$

$$= \sqrt{\frac{15.875 - \frac{(8.5)^2}{5}}{5}} = \sqrt{\frac{15.875 - 14.45}{5}} = 0.534$$

$$= \sqrt{10.984615} = 3.314$$

2- حساب الانحراف المعياري لوزن الدجاجة بعد تطبيق البرنامج .

كل دجاجة بعد تطبيق البرنامج، من المتوقع أن تزيد 0.5 كيلوجرام ، وهذا معناه أن الوزن بعد البرنامج هو : $y = x + 0.5$ ، ويكون الانحراف المعياري للوزن الجديد مساويا أيضا للانحراف المعياري للقيم الأصلية ، أى أن :

$$s_y = s_x = 0.534$$

الانحراف المعياري للوزن بعد تطبيق البرنامج يساوي 0.534 كيلوجرام .

- ثالثا : إذا ضرب كل قيمة من قيم المفردات في مقدار ثابت ، فإن الانحراف المعياري للقيم الجديدة ، يساوي الانحراف المعياري للقيم الأصلية مضروبا في الثابت ، أى أن إذا كان قيم x هي القيم الأصلية ، وكانت القيم الجديدة هي : $y = ax$ ، حيث أن a مقدار ثابت ، فإن : $s_y = as_x$.

ومثال على ذلك ، إذا كان الانحراف المعياري لدرجات عينة من الطلاب هي 4 درجات ، وإذا كان التصحيح من 50 درجة ، ويراد تعديل الدرجة ليكون التصحيح من 100 درجة، ومعنى يتم ضرب كل درجة من الدرجات الأصلية في 2 ، ومن ثم يحسب الانحراف المعياري للدرجات المعدلة كالتالي .

$$y = 2x$$
$$s_y = 2s_x = 2(4) = 8$$

إذا الانحراف المعياري للدرجات المعدلة 8 درجات .

- رابعا: إذا كان لدينا التوليفة الخطية : $y = ax + b$ ، فإن الانحراف المعياري للمتغير y هو

أيضا : $s_y = a s_x$ ، وفي المثال السابق ، لو أضاف المصحح لكل طالب 5 درجات بعد تعـــديل الدرجة من 100 ، أى أن الدرجة الجديدة هي : $y = 2x + 5$ ، فإن الانحراف المعياري هو :

$$y = 2x + 5$$

$$s_y = 2s_x = 2(4) = 8$$

مزايا وعيوب الانحراف المعياري

من مزايا الانحراف المعياري

1- أنه أكثر مقاييس التشتت استخداما .

2- يسهل التعامل معه رياضيا .

3- يأخذ كل القيم في الاعتبار .

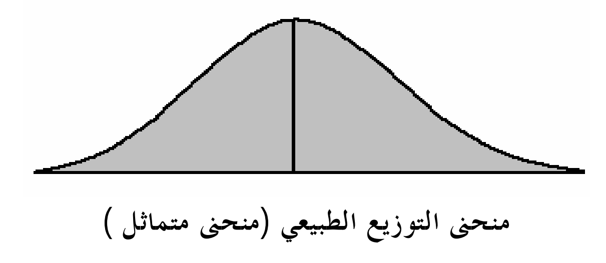
ومن عيوبه ، أنه يتأثر بالقيم الشاذة .

الفصـــــــل الخامس

بعض المقاييس الأخرى لوصف البيانات

1/5 مقدمـــــة

عند تمثيل بيانات الظاهرة في شكل منحني تكراري ، فإن هذا المنحني يأخذ أشكالا مختلفة ، فقد يكون هذا المنحنى متماثل بمعنى أن له قمة في المنتصف ، ولو أسقطنا عمودا من قمته على المحور الأفقي لشطره نصفين متماثلين ، مثل منحنى التوزيع الطبيعي ، كما هو مبين بالشكل التالي .



منحنى التوزيع الطبيعي (منحنى متماثل)

وعندما يكون الشكل متماثل ، فإن الوسط والوسيط والمنوال كلهم يقعون على نقطة واحدة ، ولكن في كثير من الحالات يكون هناك قيم كبيرة في البيانات تجذب إليها الوسط الحسابي ، وهذا معناه أن المنحنى التكراري سوف يكون له ذيل جهة اليمين ، مشيرا بوجود التواء جهة الـــيمين ، وكـــذلك العكس لو أن البيانات بها قيم صغيرة ، فإنها تجذب الوسط إليها، ويدل المنحني التكراري على وجـــود التواء جهة اليسار، كما يمكن من خلال الشكل البياني معرفة ما إذا كان توزيع البيانات منبـــسط، أو مدبب، وهذا من الناحية البيانية، إلا أن هناك مقاييس كثيرة لوصف البيانات تعتمد في حسابها علـــى مقاييس النزعة المركزية والتشتت معا، ومنها مقاييس الالتواء، والتفرطح، وبعض المقـــاييس الأخـــرى سوف يتم عرضها فيما بعد.

2/5 مقاييس الالتواء Skewness

هناك طرق كثيرة لقياس الالتواء ومنها ما يلي:

1/2/5 طريقة "بيرسون Person" في قياس الالتواء

تأخذ هذه الطريقة في الاعتبار العلاقة بين الوسط والوسيط والمنوال، في حالة ما إذا كان التوزيع قريب من التماثل وليس شديد الالتواء ، وهذه العلاقة هي: (5-1)

$$\text{المنوال} = \text{الوسط الحسابي} - ٣(\text{الوسط الحسابي} - \text{الوسيط}) \qquad (٥-١)$$

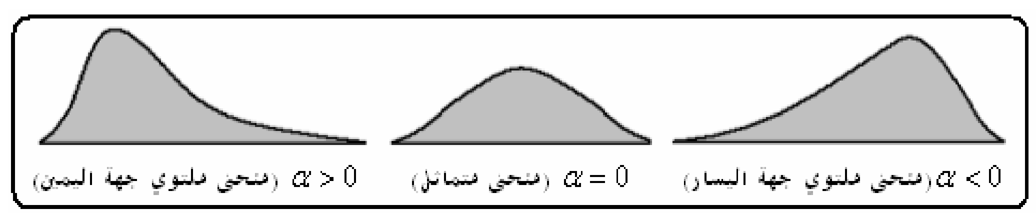
ومن ثم فإن طريقة "بيرسون" في قياس الالتواء ، تتحدد بالمعادلة التالية.

$$\alpha = \frac{3(Mean - Median)}{Standard\ Deviation} = \frac{3(\overline{x} - Med)}{S} \qquad (٥-٢)$$

حيث أن a (ألفا) هو معامل الالتواء "لبيرسون"، $\overline{x}$ الوسط الحسابي، Med هو الوسيط، S هو الانحراف المعياري، ويمكن من خلال الإشارة التي يأخذها هذا المعامل الحكم على شكل الالتواء، كما يلي :

- إذا كان (الوسط الحسابي = الوسيط) كان قيمة المعامل $(a = 0)$، ويدل ذلك على أن منحنى التوزيع التكراري متماثل.
- إذا كان (الوسط الحسابي > الوسيط) كان قيمة المعامل $(a > 0)$، ويدل ذلك على أن منحنى التوزيع التكراري ملتوي جهة اليمين .
- إذا كان (الوسط الحسابي < الوسيط) كان قيمة المعامل $(a < 0)$، ويدل ذلك على أن منحنى التوزيع التكراري ملتوي جهة اليسار.

شكل (5-1)
أشكال التواء البيانات



2/2/5 طريقة "المئين" في قياس الالتواء

المئين ينتج من ترتيب البيانات تصاعديا، ثم تقسيمها البيانات إلى 100 جزء، يفصل بينها قيم تسمى المئين، وعلى سبيل المثال يعرف المئين 15 ويرمز له بالرمز (v_{15}) على أنه القيمة التي يقل عنها 15% من القيم، ولحساب قيمة المئين p ، ونرمز له بالرمز (v_p)، يتبع نفس الفكرة المستخدمة في حساب الربيع كما يلي:

- ترتب القيم تصاعديا: $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$
- رتبة المئين: $R = (n+1)\left(\frac{p}{100}\right)$.
- إذا كانت الرتبة R عدد صحيح فإن $(v_{15} = x_{(R)})$.
- أما إذا كانت الرتبة R عدد كسري فإن قيمة المئين (v_p) تحسب بالمعادلة التالية:

$$v_p = x_{(l)} + (R - l)(x_{(u)} - x_{(l)}) \qquad (٥-٣)$$

وتعتمد فكرة المئين في قياس الالتواء على مدى قرب المئين v_p، والمئين v_{100-p}، من المئين v_{50} ، وكمثال على ذلك ، عند قياس الالتواء باستخدام المئين 20 ، والمئين 80 ، يلاحظ على الرسم التالي

حالات الالتواء :

شكل (5-2)



ومن الشكل أعلاه يلاحظ الآتي :

- إذا كان بعد المئين (v_{80}) عن المئين (v_{50}) يساوي بعد المئين (v_{20}) عن المئين (v_{50}) كـان التوزيع متماثلا .
- إذا كان بعد المئين (v_{80}) عن المئين (v_{50}) أكبر من بعد المئين (v_{20}) عن المئين (v_{50}) كـان التوزيع موجب الالتواء .
- إذا كان بعد المئين (v_{80}) عن المئين (v_{50}) أقل من بعد المئين (v_{20}) عن المئين (v_{50}) كـان التوزيع سالب الالتواء .

وبشكل عام يمكن الحكم على شكل التوزيع باستخدام معامل الالتواء المئيني، ويأخذ المعادلة التالية.

$$\alpha_{p,100-p} = \frac{(v_{100-p} - v_{50}) - (v_{50} - v_p)}{v_{100-p} - v_p} \qquad (٥-٤)$$

حيث أن : $v_p < v_{50} < v_{100-p}$ ويفضل استخدام هذا المعامل في حالة البيانات التي تحتوي علـى قيم شاذة ، وأيضا البيانات التي لا نعرف لها توزيع محدد، وعندما نستخدم المئين 25 $(v_{25} = Q_1)$، المئين 75 $(v_{75} = Q_3)$ نحصل على معامل الالتواء الربيعي ، وهو :

$$v_q = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)} \qquad (٥-٥)$$

مثـــال (5-1)

كانت درجات 8 طلاب في الاختبار النهائي في مقرر **122** إحص ، كالتالي.

66 85 52 78 80 91 74 58

والمطلوب : **1**- حساب معامل الالتواء بطريقة " بيرسون " .

2- حساب معامل الالتواء الربيعي .

الحـــــل

1- حساب معامل الالتواء بطريقة "بيرسون" .

في هذه الحالة يتم تطبيق المعادلة رقم (**2-5**) كما يلي:

- حساب الوسط الحسابي ، والانحراف المعياري :

الدرجة x	x^2
66	4356
85	7225
52	2704
78	6084
80	6400
91	8281
74	5476
58	3364
584	43890

$$\sum x = 584 \ , \ \sum x^2 = 43890$$

ويكون :

$$\bar{x} = \frac{\sum x}{n} = \frac{584}{8} = 73$$

$$s = \sqrt{\frac{\sum x^2 - (\sum x)^2 / n}{n-1}} = \sqrt{\frac{43890 - (584)^2 / 8}{8-1}}$$

$$= \sqrt{\frac{1258}{7}} = \sqrt{179.71428} = 13.406$$

- حساب الوسيط :

موقع الوسيط : *(n+1)/2=(8+1)/2=4.5*

52	58	66	74	78	80	85	91
1	2	3	4	5	6	7	8

2.25 4.5 6.75

$$Med = 74 + 0.5(78-74) = 76$$

- معامل الالتواء "بيرسون"

$$s.c = \frac{3(\bar{x} - Med)}{S} = \frac{3(73-76)}{13.406} = -0.67$$

إذا منحنى توزيع درجات الطلاب ملتوي جهة اليسار .

2- معامل الالتواء الربيعي .

لحساب معامل الالتواء الربيعي ، يتم تطبيق المعادلة رقم (**5-5**).

- حساب الربيع الأدنى .

موقع الرباعي : *(n+1)/4=(8+1)(1/4)=2.25*

إذا

$$Q_1 = 58 + (2.25 - 2)(66 - 58) = 60$$

- **حساب الرباعي الأعلى.**

موقع الرباعي : *(n+1)/(3/4)=(8+1) (3/4)=6.75*

إذا

$$Q_3 = 80 + (6.75 - 6)(85 - 80) = 83.75$$

- **الوسيط (الربيع الثاني)**

$$Med(Q_2) = 76$$

إذا معامل الالتواء الربيعي هو :

$$a_q = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)} = \frac{(83.75 - 76) - (76 - 60)}{(83.75 - 60)}$$

$$= \frac{-8.25}{23.75} = -0.35$$

إذا توزيع درجات الطلاب ملتوي جهة اليسار.

3/5 التفرطح Kurtosis

عند تمثيل التوزيع التكراري في شكل منحنى تكراري ، قد يكون هذا المـــنحنى منبـــسط ، أو مدبب ، فعندما يتركز عدد أكبر من القيم بالقرب من منتصف المنحنى، ويقل في طرفيه، يكون المنحنى مدببا ، وعندما يتركز عدد أكبر على طرفي المنحنى ، ويقل بالقرب من المنتصف يكون المنحنى مفرطحا ، أو منبسطا، ويظهر ذلك من الشكل التالي :



منحنى مدبب   **منحنى مفرطح**

ويمكن قياس التفرطح باستخدام عدد من الطرق، ومنها طريقة العزوم ، حيث يحسب معامل التفرطح (K) بتطبيق المعادلة التالية :

$$k = \frac{\frac{1}{n}\sum (x-\bar{x})^4}{s^4} \qquad (٥-٦)$$

حيث أن المقدار $\sum (x-\bar{x})^4 / n$ هو العزم الرابع حول الوسـط ، s هـو الانحـراف المعياري . ومعامل التفرطح في التوزيع الطبيعي يساوي 3 ، ومن ثم يمكن وصف منحنى التوزيع مـن حيث التفرطح ، والتدبب كما يلي :

- إذا كان $k=3$ كان منحنى التوزيع معتدلا .
- إذا كان $k>3$ كان منحنى التوزيع مدببا .
- إذا كان $k<3$ كان منحنى التوزيع منبسطا (مفرطحا) .

وبالتطبيق على بيانات المثال رقم (5-1) نجد أن: $\bar{x} = 73$

x	66	85	52	78	80	91	74	58	584
$(x-\bar{x})$	-7	12	-21	5	7	18	1	-15	0
$(x-\bar{x})^2$	49	144	441	25	49	324	1	225	1258
$(x-\bar{x})^4$	2401	20736	194481	625	2401	104976	1	50625	376246

ومن البيانات أعلاه نجد أن:

$$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{1258}{7}} = 13.406$$

$$\frac{1}{n}\sum (x-\bar{x})^2 = \frac{1}{8}(376246) = 47030.75$$

إذا معامل التفرطح هو:

$$K = \frac{47030.75}{(13.406)^4} = \frac{47030.75}{(32299.58)} = 1.456$$

إذا شكل توزيع بيانات الدرجات مفرطح.

4/5 مقاييس أخرى لوصف البيانات

هناك مقاييس أخرى يمكن استخدامها في وصف البيانات ، من حيث درجة تشتت البيانات، ومدى انتشارها، ومن هذه المقاييس ، ما يلي :

1/4/5 معامل الاختلاف Variation Coefficient

أحد مقاييس المستخدمة لقياس درجة التشتت، وفيه يحسب قيمة التشتت كنسبة مئوية من قيمة مقياس النزعة المركزية ، ومن ثم يفضل استخدام معامل الاختلاف عند مقارنة درجة تشتت بيانات مجموعتين أو أكثر مختلفة لها وحدات قياس مختلفة، بدلا من الانحراف المعياري ، أو الانحراف الربيعي ،

لأن معامل الاختلاف يعتمد على التغيرات النسبية في القيم عن مقياس النزعة المركزية، بينما يعتمد الانحراف المعياري أو الانحراف الربيعي على التغيرات المطلقة للقيم، فعند مقارنة درجة تشتت بيانات الأطوال بالسنتمتر، وبيانات الأوزان بالكيلوجرام، لا يمكن الاعتماد على الانحراف المعياري في هذه المقارنة، وإنما يستخدم معامل الاختلاف، ومن ثم يطلق عليه بمعامل الاختلاف النسبي، وفيما يلي بعض هذه المعاملات.

- **معامل الاختلاف النسبي**

ويحسب معامل الاختلاف النسبي بتطبيق المعادلة التالية:

$$v.c = \frac{s}{\bar{x}} \times 100 \qquad (٥-٧)$$

- **معامل الاختلاف الربيعي**

ويحسب هذا المعامل بتطبيق المعادلة التالية:

$$v.c_q = \frac{(Q_3 - Q_1)/2}{Med} \times 100 \qquad (٥-٨)$$

مثال (5-2)

تم اختيار مجموعتين من الأغنام النامية في أحد المزارع، وتم استخدام عليقة معينة لتسمين المجموعة الأولى، بينما تم استخدام عليقة أخرى لتسمين المجموعة الثانية ، وبعد فترة زمنية تم جمع بيانات عن أوزان المجموعتين بالكيلوجرام ، وتم الحصول على المقاييس التالية .

المقاييس	المجموعة الأولى	المجموعة الثانية
$\bar{x} =$	173	198
$s =$	23	25

والمطلوب مقارنة درجة تشتت المجموعتين:

الحـــــــل:

- معامل الاختلاف النسبي للمجموعة الأولى:

$$v.c_1 = \frac{s}{\bar{x}} \times 100 = \frac{23}{173} \times 100 = 13.3\%$$

- معامل الاختلاف النسبي للمجموعة الثانية:

$$v.c_2 = \frac{s}{\bar{x}} \times 100 = \frac{25}{195} \times 100 = 12.8\%$$

يلاحظ أن درجة تشتت أوزان المجموعة الثانية أقل من درجة تشتت أوزان المجموعة الأولى.

2/4/5 تقدير مدى الانحراف المعياري

يمكن قياس درجة تشتت البيانات من خلال تقدير المدى الذي يقع داخله الانحراف المعياري وهو:

$$\frac{\text{المدى}}{6} < \text{الانحراف المعياري} < \frac{\text{المدى}}{4}$$

$$\frac{Rang}{6} < \hat{S} < \frac{Rang}{4} \qquad (٥-٩)$$

وإذا كان المدى الذي يقع فيه الانحراف المعياري صغير دل ذلك على أن تشتت البيانات صغير، أما إذا كان المدى كبير دل ذلك على وجود تشتت كبير في البيانات، وإذا وقع الانحراف المعياري خارج المدى دل ذلك على وجود قيم شاذة.

3/4/5 الدرجة المعيارية Standardized degree

تقيس الدرجة المعيارية لقيمة معينة عدد وحدات الانحراف المعياري التي تزيد بها تقل بها هذه القيمة عن الوسط الحسابي، فإذا كان $X_1, X_2, \ldots, X_n$ ، هي قيم المشاهدات، وعددها n، وكان $\overline{X}$ هو الوسط الحسابي لهذه القيم، S هو الانحراف المعياري، فإن الدرجة المعيارية للقيمة X ، ويرمز لها بالرمز Z ، تحسب باستخدام المعادلة التالية:

$$Z = \frac{x - \overline{x}}{s} \qquad (٥-١٠)$$

ويمكن استخدام هذه الدرجة في مقارنة قيمتين أو أكثر مختلفة من حيث وحدات القياس.

مثال (5-3)

في المثال (5-2) السابق إذا تم اختيار أحد الأغنام من المجموعة الأولى بعد تطبيق البرنامج ، ووجد أن وزنه 178 كيلوجرام، وبالمثل أحد الأغنام من المجموعة الثانية، ووجد أن وزنه 180 كيلوجرام ، قارن بين هذين القيمتين من حيث أهمية كل منها في المجموعة التي تنتمي إليها.

الحل

البيانات المتاحة عن كل من المجموعتين هي:

	المجموعة الأولى	المجموعة الثانية
$\overline{X} =$	173	198
$S =$	23	25

القيمة.	178	180

للمقارنة بين الوحدتين من حيث أهمية وزن كل منها في المجموعة التي تنتمي إليها، يتم حساب الدرجة المعيارية لوزن كل منها، بتطبيق المعادلة (**5-10**).

- الدرجة المعيارية لوزن الوحدة المسحوبة من المجموعة الأولى (178 Kg.) هي:

$$z = \frac{x - \overline{x}}{s} = \frac{178 - 173}{23} = 0.22$$

- الدرجة المعيارية لوزن الوحدة المسحوبة من المجموعة الثانية (180 Kg.) هي:

$$z = \frac{x - \overline{x}}{s} = \frac{180 - 198}{25} = -0.75$$

- نجد أن الوزن 178 كيلوجرام يزيد عن الوسط الحسابي بـ 0.22 انحراف معياري ، بينما نجد أن الوزن 180 كيلوجرام يقل عن الوسط الحسابي بـ 0.75 انحراف معياري . ومن ثم الوزن الأول أهميته النسبية أعلى من الوزن الثاني.

4/4/5 القاعدة العملية

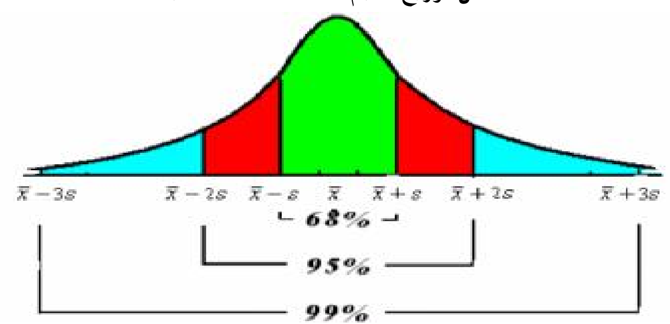
إذا كان لدينا المشاهدات التالية: $x_1, x_2, ..., x_n$ ، وكان $\overline{x}$ هو الوسط الحسابي لهذه المشاهدات، S هو الانحراف المعياري لها ، يكون منحنى توزيع هذه المشاهدات متماثل، إذا تحقق الآتي:

- 68% تقريبا من قيم هذه المشاهدات تتراوح بين $\overline{x} \pm S$.
- 95% تقريبا من قيم هذه المشاهدات تتراوح بين $\overline{x} \pm 2S$.
- 99% تقريبا من قيم هذه المشاهدات تتراوح بين $\overline{x} \pm 3S$.

ويمكن بيان ذلك من الشكل التالي:

شكل (**5-3**)

شكل توزيع القيم طبقا للقاعدة العملية



5/4/4 القاعدة النظرية

تسمى هذه القاعدة بقاعدة "تشيبشيف" ، وفكرة هذه القاعدة: في أى توزيع من التوزيعات النظرية ، فإنه على الأقل $\left(1-1/k^2\right)$ % من قيم المشاهدات تقع في المدى $\overline{x} \pm kS$ ، $k > 1$.

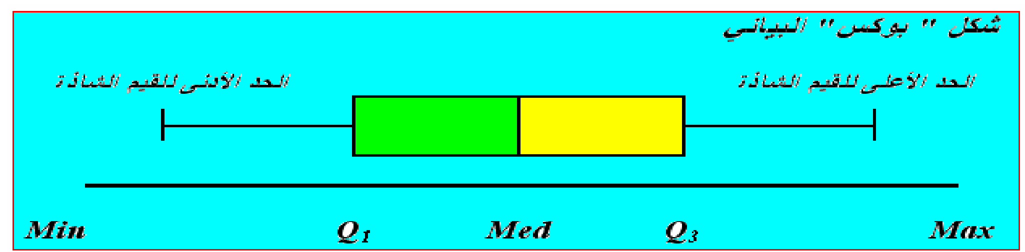
وطبقا لهذه القاعدة، فإنه على الأقل 75% من قيم المشاهدات تقع في المدى $\overline{x} \pm 2S$ ، على الأقل 89% من قيم المشاهدات تقع في المدى $\overline{x} \pm 3S$.

6/4/5 شكل "بوكس" Box Plot

شكل "بوكس" البياني هو صندوق يشبه المستطيل، بداية حافته اليسرى هو الربيع الأول Q_1 ونهاية حافته اليمنى هو الربيع الثالث Q_3 ، ويقسم الربيع الثاني (الوسيط) Med المستطيل إلى جزأين، ويخرج من كل حافة من حافتيه شعيرة، والشكل التالي يبين رسمة "بوكس" البياني:

شكل (5-4)

رسمة بوكس البياني

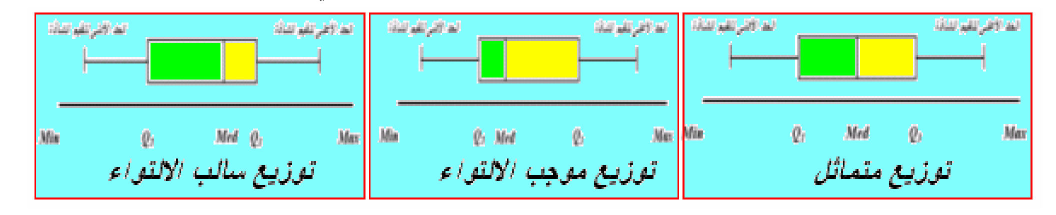


ويمكن استخدام شكل "بوكس" البياني، أعلاه في وصف البيانات من حيث الآتي:

1- من حيث التماثل: إذا كان الوسيط Med يقع في المنتصف على بعد متساوي من الرباعيين Q_3 , Q_1 كان التوزيع متماثلا ، وإذا كان الوسيط Med أقرب إلى الرباعي الأول Q_1 من الرباعي الثالث Q_3 كان التوزيع موجب الالتواء ، وأما إذا كان الوسيط Med أقرب إلى الرباعي الثالث Q_3 من الرباعي الأول Q_1 كان التوزيع سالب الالتواء . ويظهر ذالك كما في الشكل التالي :

شكل (5-5)

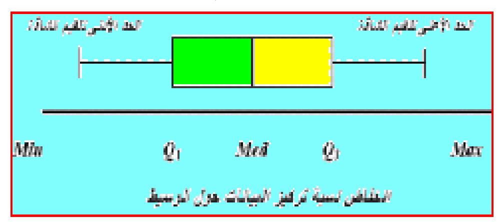
وصف شكل الالتواء باستخدام رسمة بوكس البياني

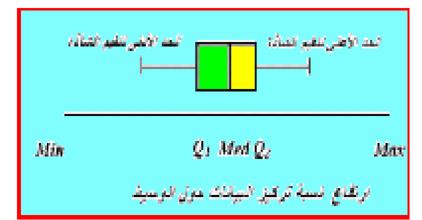


2- من حيث تركز البيانات: إذا كان الصندوق Box ضيق دل ذلك على تركز نسبة كبيرة من البيانات حول الوسيط، وإذا كان الصندوق واسع دل ذلك على انخفاض نسبة تركز البيانات حول الوسيط، والشكل التالي يبين ذلك.

شكل (5-6)

وصف درجة تركز البيانات باستخدام رسمة بوكس البياني





3- من حيث وجود قيم شاذة: إذا وقعت قيم بعض المشاهدات خارج الحدين الأدنى والأعلى الشاذ ،

كانت هذه القيم شاذة ، وتظهر هذه القيم على الرسم في شكل نجوم (*) ، والشكل التالي يبين طريقة عرض القيم الشاذة الدنيا والعليا على الرسم .

شكل (5-7)

تحديد القيم الشاذة باستخدام رسمة بوكس البياني



طريقة حساب حدي القيم الشاذة

لحساب الحدين الأعلى والأدنى للقيم الشاذة، يتبع الخطوات التالية:

- حساب الانحراف الرباعي: $Q = (Q_3 - Q_1)/2$
- حساب الحد الأدنى للقيم الشاذة (Low)، وهو: $Low = Q_1 - 3Q$
- حساب الحد الأعلى للقيم الشاذة (Upp) ، وهو: $UPP = Q_3 + 3Q$

وإذا وقعت قيم خارج الحدين تعتبر هذه القيم من القيم الشاذة.

مثـــــــال(5-4)

فيما يلي الإنفاق الاستهلاكي بالألف ريال خلال الشهر لعينة حجمها 12 أسرة:

7 2 8 11 6 5 10 9 3 18 10 6

والمطلوب:

1- رسم شكل "بوكس" البياني

2- اكتب تحليل وصفي لهذه البيانات.

الحـــــــــل

1- رسم شكل " بوكس البياني "

- ترتيب القيم تصاعديا .

2 3 5 6 6 7 8 9 10 10 11 18

- تحديد أقل وأعلى إنفاق استهلاكي، وحساب الرباعيات:

$Min = 2$ $Max = 18$

الرباعي الأدنى Q_1 :

موقع الرباعي $(n+1)(1/4)=(13/4)=3.25$

إذا قيمة Q_1 هي: $Q_1 = 5 + 0.25(6-5) = 5.25$

الوسيط Med:

موقع الوسيط $(n+1)(1/2)=(13/2)=6.5$

إذا قيمة Med هي: $Med = 7 + 0.5(8-7) = 7.5$

الرباعي الأعلى Q_3:

موقع الرباعي $(n+1)(3/4)=(13)(3/4)=9.75$

إذا قيمة Q_3 هي: $Q_3 = 10 + 0.75(10-10) = 10$

- حساب الحدين الأعلى والأدنى الشاذ .

الانحراف الربيعي: $Q = (10-5.25)/2 = 2.375$

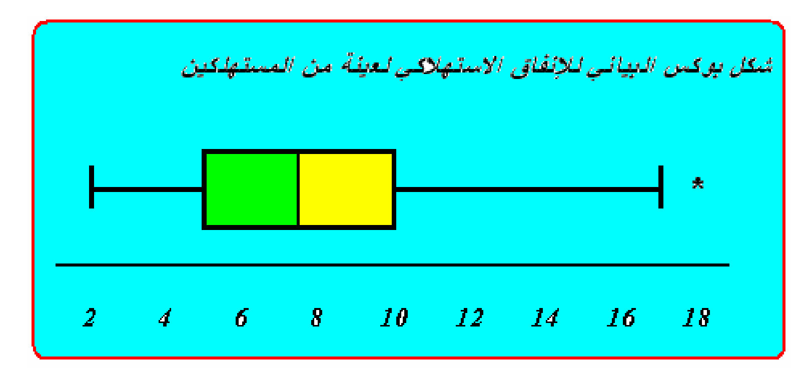
الحد الأدنى للقيم الشاذة:

$$Low = Q_1 - 3Q = 5.25 - 3(2.375) = -1.875$$

الحد الأعلى للقيم الشاذة :

$$Upp = Q_3 + 3Q = 10 + 3(2.375) = 17.125$$

- رسم شكل "بوكس"



2- تحليل وصفي من خلال الشكل أعلاه:

- درجة التماثل : التوزيع قريب جدا من التماثل لوقوع الوسيط في المنتصف .
- تركز البيانات: حوالي 60% من القيم تتركز حول الوسيط.
- القيم الشاذة: توجد قيمة شاذة عليا هي القيمة 18.

ويمكن استخدام شكل "بوكس" البياني لمقارنة مجموعتين أو أكثر .

الفصـــــــــــل السادس
الارتباط والانحدار الخطي البسيط

1/6 مقـــــــــــدمة

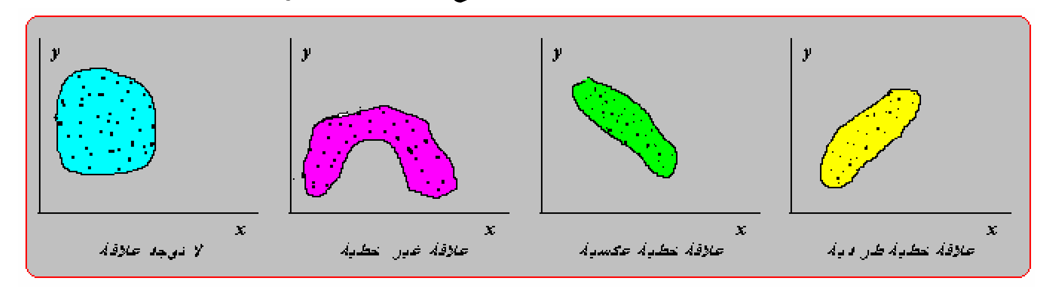
في الفصول الثلاث السابقة تم عرض بعض المقاييس الوصفية، مثل مقاييس النزعة المركزية، والتشتت، ومقاييس الالتواء والتفرطح، وغيرها من المقاييس الأخرى والتي يمكن من خلالها وصف شكل توزيع البيانات التي تم جمعها عن متغير واحد، وننتقل من التعامل مع متغير واحد إلى التعامل مع متغيرين أو أكثر، ويتناول هذا الفصل دراسة وتحليل العلاقة بين متغيرين، وذلك باستخدام بعض طرق التحليل الإحصائي مثل تحليل الارتباط، والانحدار الخطي البسيط، فإذا كان اهتمام الباحث هو دراسة العلاقة بين متغيرين استخدم لذلك أسلوب تحليل الارتباط، وإذا كان اهتمامه بدراسة أثر أحد المتغيرين على الآخر استخدم لذلك أسلوب تحليل الانحدار، ومن الأمثلة على ذلك:

1- الإنفاق، والدخل العائلي.
2- سعر السلعة، والكمية المطلوبة منها.
3- الفترة الزمنية لتخزين الخبز، وعمق طراوة الخبز.
4- تقديرات الطلاب في مقرر الإحصاء، وتقديراهم في مقرر الرياضيات.
5- كميات السماد المستخدمة، وكمية الإنتاج من محصول معين تم تسميده بهذا النوع من السماد.
6- عدد مرات ممارسة نوع معين من الرياضة البدنية، ومستوى الكلسترول في الدم.
7- وزن الجسم، وضغط الدم.

والأمثلة على ذلك في المجال التطبيقي كثيرة، فإذا كان لدينا المتغيرين $(x \ , \ y)$ ، وتم جمع بيانات عن أزواج قيم هذين المتغيرين، وتم تمثيلها بيانيا فيما يسمى بشكل الانتشار، فإن العلاقة بينها تأخذ أشكالا مختلفة على النحو التالي :

شكل(6-1)

شكل الانتشار لبيان نوع العلاقة بين x , y



2/6 الارتباط الخطى البسيط Simple Correlation

إذا كان الغرض من التحليل هو تحديد نوع وقوة العلاقة بين متغيرين ، يستخدم تحليل الارتباط ، وأما إذا كان الغرض هو دراسة وتحليل أثر أحد المتغيرين على الآخر ، يستخدم تحليل الانحدار، وفي

هذا الفصل يتم عرض أسلوب تحليل الارتباط الخطي البسيط، أي في حالة افتراض أن العلاقة بين المتغيرين تأخذ الشكل الخطي ، وسوف يجرى حسابه في حالة البيانات الكمية ، والبيانات الوصفية المقاسة بمعيار ترتيبي.

1/2/6 الغرض من تحليل الارتباط الخطى البسيط

الغرض من تحليل الارتباط الخطي البسيط هو تحديد نوع وقوة العلاقة بين متغيرين، ويرمز له في حالة المجتمع بالرمز r (رو)، وفي حالة العينة بالرمز r، وحيث أننا في كثير من النواحي التطبيقية نتعامل مع بيانات عينة مسحوبة من المجتمع، سوف نهتم بحساب معامل الارتباط في العينة r كتقدير لمعامل الارتباط في المجتمع، ومن التحديد السابق للغرض من معامل الارتباط، نجد أنه يركز على نقطتين هما:

- **نوع العلاقة:ـــ** وتأخذ ثلاث أنواع حسب إشارة معامل الارتباط كما يلي:

1- إذا كانت إشارة معامل الارتباط سالبة $(r < 0)$ توجد علاقة عكسية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه انخفاض في المتغير الثاني، والعكس.

2- إذا كانت إشارة معامل الارتباط موجبة $(r > 0)$توجد علاقة طردية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه زيادة في المتغير الثاني، والعكس .

3- إذا كان معامل الارتباط قيمته صفرا $(r = 0)$ دل ذلك على انعدام العلاقة بين المتغيرين.

- **قوة العلاقة:ـــ** ويمكن الحكم على قوة العلاقة من حيث درجة قربها أو بعدها عن (± 1)، حيث أن قيمة معامل الارتباط تقع في المدى$(-1 < r < 1)$، وقد صنف بعض الإحصائيين درجات لقوة العلاقة يمكن تمثيلها على الشكل التالي:

شكل (6-2)

درجات قوة معامل الارتباط

ارتباط عكسي					ارتباط طردي				
قوي جدا	قوي	متوسط	ضعيف	ضعيف جدا	ضعيف جدا	ضعيف	متوسط	قوي	قوي جدا
-1 ... -0.9	-0.9 ... -0.7	-0.7 ... -0.5	-0.5 ... -0.3	-0.3 ... 0	0 ... 0.3	0.3 ... 0.5	0.5 ... 0.7	0.7 ... 0.9	0.9 ... 1

تام (-1) — منعدم (0) — تام (1)

2/2/6 معامل الارتباط الخطى البسيط " لبيرسون" Pearson

في حالة جمع بيانات عن متغيرين كميين (y , x)، يمكن قياس الارتباط بينهما، باستخدام طريقة "بيرسون" Pearson، ومن الأمثلة على ذلك: قياس العلاقة بين الوزن والطول، والعلاقة بين الإنتاج والتكلفة، والعلاقة بين الإنفاق الاستهلاكي والدخل، والعلاقة بن الدرجة التي حصل عليها الطالب وعدد ساعات الاستذكار، وهكذا الأمثلة على ذلك كثيرة.

ولحساب معامل الارتباط في العينة ، تستخدم صيغة " بيرسون" التالية :

$$r=\frac{S_{xy}}{S_x\,S_y}=\frac{\dfrac{\sum(x-\bar{x})(y-\bar{y})}{(n-1)}}{\sqrt{\dfrac{\sum(x-\bar{x})^2}{(n-1)}}\sqrt{\dfrac{\sum(y-\bar{y})^2}{(n-1)}}} \qquad (٦-١)$$

حيث أن :

$S_{xy}=\sum(x-\bar{x})(y-\bar{y})/(n-1)$: هو التغاير بين $(x\ ,\ y)$،

$S_x=\sqrt{\sum(x-\bar{x})^2/(n-1)}$: هو الانحراف المعياري لقيم (x)،

$S_y=\sqrt{\sum(y-\bar{y})^2/(n-1)}$: هو الانحراف المعياري لقيم (y).

ويمكن اختصار الصيغة السابقة على النحو التالي:

$$r=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}} \qquad (٦-٢)$$

مثـــــال (1-6)

فيما يلي المساحة المتزرعة بالأعلاف الخضراء بالألف هكتار، وإجمالي إنتاج اللحوم بالألف طن، خلال الفترة من 1995حتى عام 2002.

السنة	1995	1996	1997	1998	1999	2000	2001	2002
المساحة	305	313	297	289	233	214	240	217
الكمية	592	603	662	607	635	699	719	747

والمطلوب: حساب معامل الارتباط بين المساحة والكمية، وما هو مدلوله ؟

الحــــل

بفرض أن (x) هي المساحة المتزرعة، (y) هي الكمية، ولحساب معامل الارتباط بين $(x\ ,\ y)$ يتم تطبيق المعادلة (6-2)، وذلك على النحو التالي:

- حساب الوسط الحسابي لكل من المساحة، والكمية $(\bar{x}\ ,\ \bar{y})$.

$$\bar{x}=\frac{\sum x}{n}=\frac{2108}{8}=263.5\ ,\ \bar{y}=\frac{\sum y}{n}=\frac{5264}{8}=658$$

- حساب المجاميع

x	y	$x-\bar{x}$	$(x-\bar{x})^2$	$y-\bar{y}$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
305	592	41.5	1722.25	-66	4356	-2739
313	603	49.5	2450.25	-55	3025	-2722.5
297	662	33.5	1122.25	4	16	134
289	607	25.5	650.25	-51	2601	-1300.5
233	635	-30.5	930.25	-23	529	701.5
214	699	-49.5	2450.25	41	1681	-2029.5
240	719	-23.5	552.25	61	3721	-1433.5
217	747	-46.5	2162.25	89	7921	-4138.5
2108	5264	0	12040	0	23850	-13528

$$\sum(x-\bar{x})^2 = 12040 \ , \ \sum(y-\bar{y})^2 = 23850 \ ,$$
$$\sum(x-\bar{x})(y-\bar{y}) = -13528$$

إذا معامل الارتباط قيمته هي:

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}} = \frac{-13528}{\sqrt{12040}\sqrt{23850}}$$
$$= \frac{-13528}{(109.727)(154.434)} = \frac{-13528}{16945.619} = -0.798$$

- يوجد ارتباط عكسي قوي بين المساحة المنزرعة، وكمية إنتاج اللحوم.

تبسيط العمليات الحسابية:

في بعض الأحيان، يكون استخدام صيغة المعادلة (**6-2**) في غاية الصعوبة، خاصة إذا لازم العمليات الحسابية قيما كسرية، من أجل ذلك يمكن تبسيط الصيغة (**6-2**) إلى صيغة أسهل تعتمد على مجموع القيم وليس على انحرافات القيم عن وسطها الحسابي، وهذه الصيغة هي:

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}} \qquad (٦-٣)$$

وبالتطبيق على بيانات المثال السابق ، يتبع الآتي :

- حساب المجاميع:

x	y	xy	x^2	y^2
305	592	180560	93025	350464
313	603	188739	97969	363609
297	662	196614	88209	438244

المجاميع المطلوبة
$\sum x = 2108 \ , \ \sum y = 5264$ $\sum xy = 1373536$

289	607	175423	83521	368449
233	635	147955	54289	403225
214	699	149586	45796	488601
240	719	172560	57600	516961
217	747	162099	47089	558009
2108	5264	1373536	567498	3487562

$\sum x^2 = 567498$

$\sum y^2 = 3487562$

- حساب معامل الارتباط:

باستخدام المجاميع السابقة، وبالتطبيق على المعادلة (**6-3**) أعلاه، نجد أن معامل الارتباط قيمته هي:

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$

$$= \frac{1373536 - \frac{(2108)(5264)}{8}}{\sqrt{\left(567498 - \frac{(2108)^2}{8}\right)\left(3487562 - \frac{(5264)^2}{8}\right)}}$$

$$= \frac{-13528}{\sqrt{(12040)(23850)}} = \frac{-13528}{16945.619} = -0.798$$

وهي نفس النتيجة السابقة:

3/2/6 معامل ارتباط الرتب (اسبيرمان) Spearman

إذا كانت الظاهرة محل الدراسة تحتوي على متغيرين وصفيين ترتيبين، ومثال على ذلك قياس العلاقة بين تقديرات الطلبة في مادتين ، أو العلاقة بين درجة تفضيل المستهلك لسلعة معينة ، ومستوى الدخل، فإنه يمكن استخدام طريقة "بيرسون" السابقة في حساب معامل ارتباط يعتمد على رتب مستويات المتغيرين كبديل للقيم الأصلية ، ويطلق على هذا المعامل " معامل ارتباط اسبيرمان " Spearman ، ويعبر عنه بالمعادلة التالية :

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} \qquad (٦-٤)$$

حيث أن d هي الفرق بين رتب مستويات المتغير الأول x ، ورتب مستويات المتغير الثاني y ، أي أن : $d = R_x - R_y$.

مثـــــــال (6-2)

فيما يلي تقديرات 10 طلاب في مادتي الإحصاء، والاقتصاد:

تقديرات إحصاء	أ	جـ+	د	د+	ب+	جـ+	أ+	ب	ب+	ب+
تقديرات اقتصاد	أ+	د	جـ	جـ	أ	ب	ب+	ب	جـ	ب

والمطلوب:

1- احسب معامل الارتباط بين تقديرات الطلبة في المقررين.

2- وما هو مدلوله ؟

الحـــــــــل

1- بفرض أن x هي تقديرات الإحصاء، y هي تقديرات الاقتصاد، يمكن حساب معامل الارتبـــاط بينهما باستخدام المعادلة (**6-4**)، وذلك بإتباع الآتي:

الرتب	1	2	3	4	5	6	7	8	9	10
تقديرات إحصاء	أ+	أ	ب+	ب+	ب+	ب	جـ+	جـ+	د+	د
رتب x	1	2	(3+4+5)/3=4			6	(7+8)/2=7.5		9	10
تقديرات اقتصاد	أ+	أ	ب+	ب	ب	ب	جـ	جـ	جـ	د
رتب y	1	2	3	(4+5+6)/3=5			(7+8+9)/3=8			10

- إذا يمكن حساب المجموع: $\sum d^2$ كما يلي:

x	y	رتب x	رتب y	d	d^2
أ	أ+	2	1	1	1
جـ+	د	7.5	10	-2.5	6.25
د	جـ	10	8	2	4
د+	جـ	9	8	1	1
ب+	أ	4	2	2	1
جـ+	ب	7.5	5	2.5	6.25
أ+	ب+	1	3	-2	4
ب	ب	6	5	1	1
ب+	جـ	4	8	-4	16
ب+	ب	4	5	-1	1
					44.5

$$\sum d^2 = 44.5$$

- معامل الارتباط هو:

$$r = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

$$= 1 - \frac{6(44.5)}{10(10^2-1)} = 1 - \frac{267}{990}$$

$$= 1 - 0.2697 = 0.7303$$

2- مدلول معامل الارتباط :

بما أن $r = 0.703$ ، ويدل ذلك على وجود ارتباط طردي قوي بين تقديرات الطالـــب في مادة الإحصاء ، ومادة الاقتصاد .

ملحوظة:- يمكن استخدام صيغة معامل ارتباط "اسبيرمان" في حساب الارتباط بين متغيرين كمـــين، حيث يتم استخدام رتب القيم التي يأخذها المتغير، ونترك للطالب القيام بحساب معامل ارتباط الرتـــب بين المساحة والكمية في مثال (**1-5**) السابق، وعليه أن يقوم بتفسير النتيجة: (معاونة : $\sum d^2 = 148$

)

3/6 الانحدار الخطي البسيط Simple Regression

إن الغرض من استخدام أسلوب تحليل الانحدار الخطي البسيط، هو دراسة وتحليل أثر متغير كمي على متغير كمي آخر، ومن الأمثلة على ذلك ما يلي:

- دراسة أثر كمية السماد على إنتاجية الدونم.
- دراسة أثر الإنتاج على التكلفة.
- دراسة أثر كمية البروتين التي يتناولها الأبقار على الزيادة في الوزن.
- أثر الدخل على الإنفاق الاستهلاكي.

وهكذا هناك أمثلة في كثير من النواحي الاقتصادية، والزراعية، والتجارية، والعلوم السلوكية، وغيرها من المجالات الأخرى.

1/3/6 نموذج الانحدار الخطي

في تحليل الانحدار البسيط، نجد أن الباحث يهتم بدراسة أثر أحد المتغيرين ويسمى بالمتغير المستقل أو المتنبأ منه، على المتغير الثاني ويسمى بالمتغير التابع أو المتنبأ به، ومن ثم يمكن عرض نموذج الانحدار الخطي في شكل معادلة خطية من الدرجة الأولى، تعكس المتغير التابع كدالة في المتغير المستقل كما يلي:

$$y = \beta_0 + \beta_1 x + e \quad (٦-٥)$$

حيث أن:

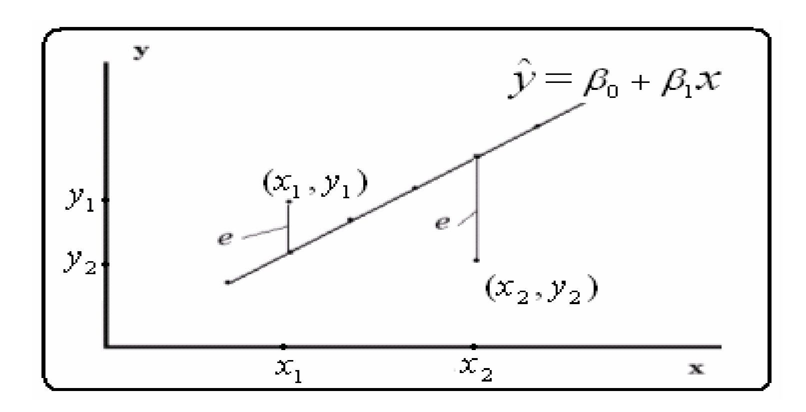
y : هو المتغير التابع (الذي يتأثر)

x : هو المتغير المستقل (الذي يؤثر)

b_0 : هو الجزء المقطوع من المحور الرأسي y ، وهو يعكس قيمة المتغير التابع في حالة انعدام قيمة المتغير المستقل x ، أي في حالة $x = 0$

b_1 : ميل الخط المستقيم $(b_0 + b_1 x)$ ، ويعكس مقدار التغير في y إذا تغيرت x بوحدة واحدة.

e : هو الخطأ العشوائي، والذي يعبر عن الفرق بين القيمة الفعلية y ، والقيمة المقدرة $\hat{y} = b_0 + b_1 x$ ، أي أن : $e = y - (b_0 + b_1 x)$ ، ويمكن توضيح هذا الخطأ على الشكل التالي لنقط الانتشار.



2/3/6 تقدير نموذج الانحدار الخطي البسيط

يمكن تقدير معاملات الانحدار $(b_1 \ , \ b_0)$ في النموذج **(6-5)** باستخدام طريقة المربعات الصغرى، وهذا التقدير هو الذي يجعل مجموع مربعات الأخطاء العشوائية $\sum e^2 = \sum \left(y-(b_0+b_1x)\right)^2$ أقل ما يمكن، ويحسب هذا التقدير بالمعادلة التالية:

$$\hat{\beta}_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} \ , \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad (٦-٦)$$

حيث أن $\bar{x}$ هو الوسط الحسابي لقيم x، $\bar{y}$ هو الوسط الحسابي لقيم y، وتكون القيمة المقدرة للمتغير التابع هو: $\hat{y} = \hat{b}_0 + \hat{b}_1 x$ ، ويطلق على هذا التقدير " تقدير معادلة انحدار y على x .

مثال (6-3)

فيما يلي بيانات عن كمية البروتين اليومي بالجرام التي يحتاجها العجل الرضيع، ومقدار الزيادة في وزن العجل بالكجم، وذلك لعينة من العجول الرضيعة حجمها 10.

كمية البروتين	10	11	14	15	20	25	46	50	59	70
الزيادة في الوزن	10	10	12	12	13	13	19	15	16	20

والمطلوب :

1- ارسم نقط الانتشار، وما هو توقعاتك لشكل العلاقة ؟

2- قدر معادلة انحدار الوزن على كمية البروتين.

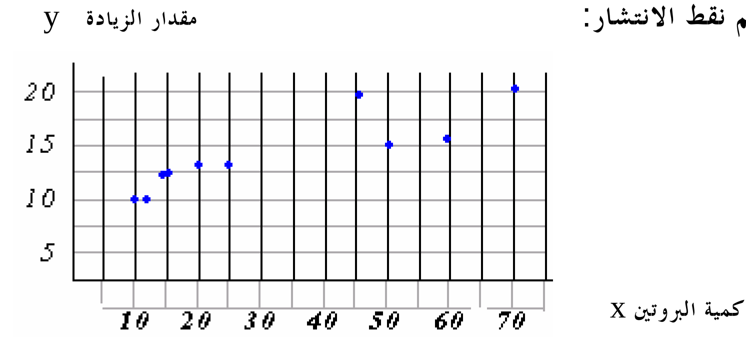
3- فسر معادلة الانحدار.

4- ما هو مقدار الزيادة في الوزن عند إعطاء العجل 50 جرام من البروتين ؟ وما هو مقدار الخطأ العشوائي؟

5- ارسم معادلة الانحدار على نقط الانتشار في المطلوب **(1)** .

الحـــــل

1- رسم نقط الانتشار:



من المتوقع أن يكون لكمية البروتين أثر طردي (إيجابي) على مقدار الزيادة في الوزن.

2- تقدير معادلة الانحدار.

بفرض أن x هي كمية البروتين، y هي مقدار الزيادة في الوزن، يمكن تطبيق المعادلتين في (**6-6**)، ومن ثم يتم حساب المجاميع التالية:

كمية البروتين x	الزيادة في الوزن y	$x\ y$	x^2
10	10	100	100
11	10	110	121
14	12	168	196
15	12	180	225
20	13	260	400
25	13	325	625
46	19	874	2116
50	15	750	2500
59	16	944	3481
70	20	1400	4900
320	140	5111	14664

المجاميع المطلوبة

$\sum x = 320$

$\sum y = 140$

$\sum xy = 5111$

$\sum x^2 = 14664$

إذا الوسط الحسابي:

$$\bar{x} = \frac{\sum x}{n} = \frac{320}{10} = 32$$

$$\bar{y} = \frac{\sum x}{n} = \frac{140}{10} = 14$$

- بتطبيق المعادلة الأولى في (**6-6**) يمكن حساب $\hat{b}_1$ كما يلي:

$$\hat{b}_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{(10)(5111) - (320)(140)}{(10)(14664) - (320)^2}$$

$$= \frac{6310}{44240} = 0.1426$$

- بتطبيق المعادلة الثانية في (**6-6**) يمكن حساب $\hat{b}_0$ كما يلي:

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x} = 14 - (0.1426)(32) = 9.4368$$

- إذا معادلة الانحدار المقدرة، هي:

$$\hat{y} = 9.44 + 0.143x$$

3- تفسير المعادلة:

- الثابت $\hat{b}_0 = 9.44$: يدل على أنه في حالة عدم استخدام البروتين قي التغذيـــة، فـــإن الوزن يزيد 9.44 كجم.
- معامل الانحدار $\hat{b}_1 = 0.143$: يدل على أنه كلما زادت كمية البروتين جرام واحـــد، حدث زيادة في وزن العجل بمقدار 0.143 كجم، أى زيادة مقدارها 143 جرام.

4- مقدار الزيادة في الوزن عند $x = 50$ هو:

$$\hat{y} = 9.44 + 0.143(50) = 16.59$$

وأما ومقدار الخطأ العشوائي هو:

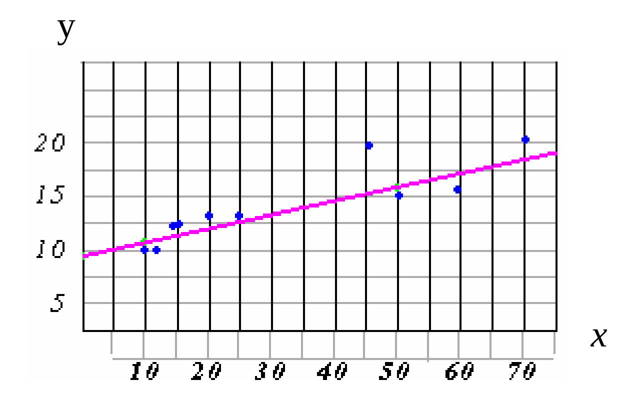
$$\hat{e}_{x=50} = y_{x=50} - \hat{y}_{x=50} = 15 - 16.59 = -1.59$$

5- رسم معادلة الانحدار على نقط الانتشار.

يمكن رسم معادلة خط مستقيم إذا علم نقطتين على الخط المستقيم.

x	50	10
$\hat{y}$	16.59	10.87

إذا معادلة الانحدار هي:



الفصـــــل السابع

الاحتمـــالات وتطبيقاتها

Probabilities and its Applications

1/7 مقـــــدمة

كلمة "احتمال" هي كلمة ينطق بها الكثير من الناس، فبعض خبراء الأرصاد الجوية يقولون من المحتمل سقوط أمطار اليوم، احتمال ارتفاع في درجات الحرارة، وبعض خبراء البورصة يقولون احتمال ارتفاع قيمة الأسهم المتداولة في سوق المال لشركة معينة، خلال هذا اليوم، واحتمال نجاح طالب، واحتمال إصابة نوع معين من الفاكهة بنوع من البكتريا، وهكذا، يكثر نطق الأفراد بها وربما يجهلون معناها. فماذا تعني كلمة احتمال؟

يقصد بهذه الكلمة فرصة حدوث أو وقوع حادثة معينة، وتستخدم الاحتمالات في كثير من النواحي التطبيقية، مثل المجالات الاقتصادية، والتجارية، والزراعية، والطبية، والسلوكية، وغيرها، خاصة عند اتخاذ القرار في دراسات الجدوى، والتنبؤ بسلوك الظواهر المختلفة، ولكي يمكن فهم موضوع الاحتمال، وأهميته في النواحي التطبيقية، نقوم بعرض بعض المفاهيم الخاصة بالاحتمالات.

2/7 بعض المفاهيم الخاصة بالاحتمال

- **التجربة العشوائية** Randomized Experiment

هي أي عملية تتم يمكن تحديد كل النتائج الممكنة لها، ولكن لا يمكن مسبقا تحديد النتيجة التي ستظهر أو تحدث، ومثال على ذلك عند إلقاء قطعة عملة معدنية مرة واحدة، فإن النتائج الممكنة لها نتيجتان هما: "ظهور الصورة" ويرمز لها بالرمز H، أو "ظهور الكتابة" ويرمز لها بالرمز T، أي أن النتائج الممكنة هي: { H , T }، وقبل إلقاء القطعة ، لا يمكن تحديد أي من النتيجتين سوف تظهر.

- **فراغ العينة** Sample Space

هي مجموعة النتائج الممكنة للتجربة، ويرمز لها بالرمز S، ويرمز لعدد النتائج المكونة لفراغ العينة بالرمز $n(S)$، ومن الأمثلة على ذلك:

1- عند إلقاء قطعة عملة غير متحيزة مرة واحدة، نجد أن فراغ العينة هو: S:{H , T } ، وعدد النتائج هي: $n(S) = 2$.

2- عند إلقاء قطعة عملة غير متحيزة مرتين (إلقاء قطعتين مرة واحدة)، فإن فراغ العينة يمكن الحصول عليه من خلال شجرة الاحتمالات كما يلي:

الرمية الأولى | الرمية الثانية | S

H — H : HH
H — T : HT
T — H : TH
T — T : TT

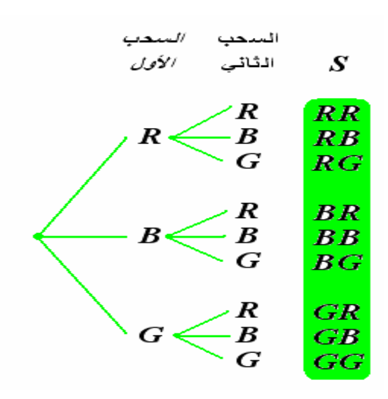
أي أن $n(S) = 4$

3- عند رمي زهرة نرد غير متحيزة مرة واحدة، فإن فراغ العينة هو مجموعة عدد النقاط التي تظهر على الوجه، وهي: S:{1, 2, 3, 4, 5, 6} ، أي أن : $n(S) = 6$.

4- عند إلقاء قطعة عملة غير متحيزة عدد من المرات حتى نحصل على الصورة مرة واحدة، نجد أن التجربة هي عدد من المحاولات يتم إيقافها عندما نحصل على الصورة مرة واحدة ، إذا فراغ العينة هو :

S:{H, TH, TTH, TTTH,.......}، ويكون : $n(S) = \infty$.

5- عند سحب كرتين بدون إرجاع من كيس به خمس كرات حمراء (red)، ثلاث كرات زرقاء (blue)، وكرتان خضراء (green)، نجد أن فراغ العينة هو:



أي أن: $n(S) = (10 \times 9) = 90$ ، (لأنها حالات غير متزنة).

6- عند فرز صندوق به خمس وحدات من سلعة معينة، يكون فراغ العينة لعدد الوحـــدات المعيبة هو .... واجب منزلي

- **الحادث** Event

هو فئة جزئية من النتائج المكونة لفراغ العينة، ويرمز للحادث بحرف من الحروف الهجائيـــة [A, B, C,...] ، وينقسم الحادث إلي نوعين هما:

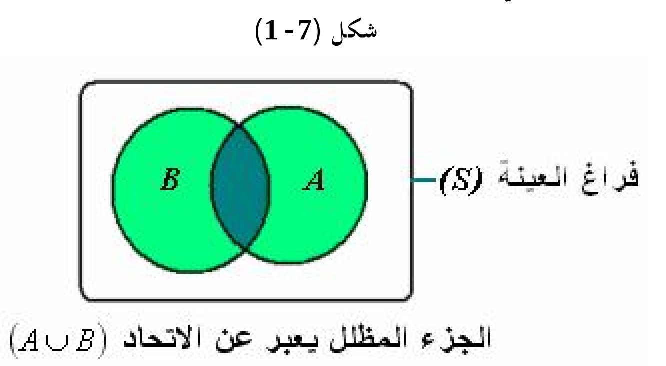
1- **حادث بسيط** Simple Event: وهو الذي يحتوي على نتيجة واحدة من النتـــائج المكونة لفراغ العينة.

2- حادث مركب Component Event: ويشمل نتيجتين أو أكثر من النتائج المكونة لفراغ العينة، أي أن الحادث المركب يمكن تقسيمه إلى حوادث بسيطة.
ويرمز لعدد النتائج المكونة للحادث بالرمز $n(A)$, $n(B)$,... وهكذا.
فعند إلقاء قطعة عملة غير متحيزة مرتين ، وعرف الحادث A بأنه ظهور الصورة مرتين ، والحادث B ظهور الصورة مرة واحدة على الأقل ، نجد أن فراغ العينة في هذه الحالة هي S:{HH, HT, TH, TT} ، وبالنسبة للحادث A فهو حادث بسيط ، يشمل نتيجة واحدة هي A:{HH} ، أي أن n(A)=1 ، أما الحادث B فهو حادث مركب يشمل ثلاث نتائج هي B:{HT, TH, HH} ، أي أن n(B)=3 ، وهذا الحادث يمكن تقسيمه إلى أحداث بسيطة .

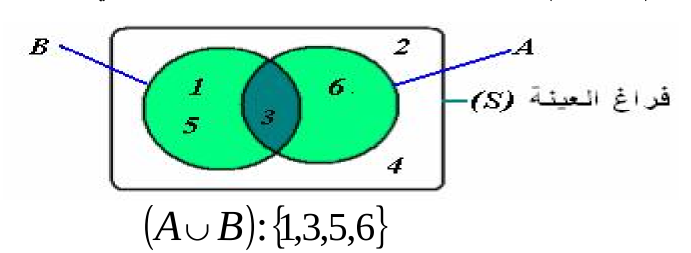
- **الاتحاد ($\cup$)** Union

يعبر اتحاد الحادثان A , B عن وقوع أحدها على الأقل، وبمعنى آخر وقوع الأول أو الثاني أو كلاهما، ويعبر عن ذلك رياضيا $(A \cup B)$ أو $(A\ or\ B)$، ويمكن الاستعانة بشكل "فن" Ven. Diagram كما يلي:

شكل (7-1)



ومثال على ذلك ، عند إلقاء زهرة نرد متزنة مرة واحدة ، وعرف الحادث A بأنه ظهور وجه يقبل القسمة على 3 ، والحادث B بأنه ظهور عدد فردي، يلاحظ أن:
S:{1,2,3,4,5,6}, A:{3,6}, B:{1,3,5} ، ويكون اتحاد الحادثان A , B هو: $(A \cup B):\{1,3,5,6\}$ ، ويعبر عن ذلك في شكل Ven كما يلي:

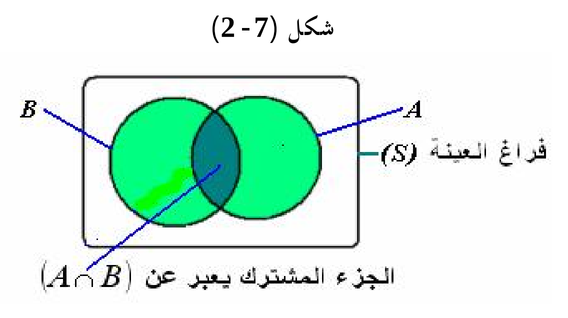


$(A \cup B):\{1,3,5,6\}$

- **التقاطع $(\cap)$** Intersection

يعبر تقاطع الحادثان A , B عن وقوع الاثنان في آن واحد ، ويشمل كل النتائج المشتركة بين الحادثين، ويعبر عن ذلك رياضيا $(A \cap B)$ أو $(A\ and\ B)$ ، ويظهر ذلك في شكل "فن" كما يلي :

شكل (7-2)

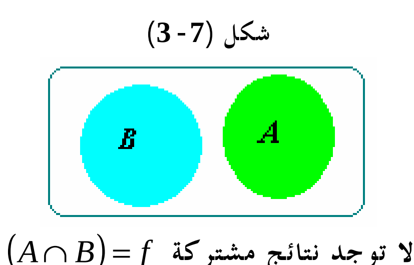


ففي المثال السابق ، نجد أن $(A \cap B):\{3\}$.

- **الأحداث المتنافية** Mutually Exclusive evens

يقال أن الحادثان A ,B متنافيان، إذا كان وقوع أحدهما ينفي وقوع الحدث الآخر، بمعنى استحالة وقوعهما في آن واحد، ومن ثم يكون نتيجة تقاطع الحادثان المتنافيان هي الفئة الخالية ويرمز لها بالرمز f أي أن $A \cap B = f$ ، ويمكن تمثيلها بشكل " فن " كما يلي:

شكل (7-3)

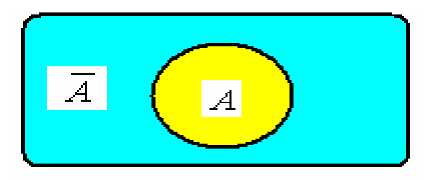


لا توجد نتائج مشتركة $(A \cap B) = f$

- **الحادث المكمل** Compliment Event

الحادث المكمل للحادث A هو الذي ينفي وقوعه، بمعنى آخر هو الحادث الذي يشمل كل نتائج التجربة باستثناء النتائج المكونة للحادث A، ويرمز للحادث المكمل بالرمز $\overline{A}$ ، ومن ثم نستنتج أن : $\left(A \cap \overline{A}\right) = f$, $\left(A \cup \overline{A}\right) = S$ كما هو مبين بالشكل التالي:

شكل (7-4)



مثـــــال (7-1)

ألقيت قطعة عملة غير متحيزة ثلاث مرات، وعرفت الأحداث التالية:

الحادث A ظهور الصورة مرتين.

الحادث B ظهور الصورة مرة واحدة.

الحادث C ظهور الصورة في الرمية الأولى.

والمطلوب:

1- إيجاد الأحداث الخاصة بالاتحاد:

$$A\cup B\ ,\ A\cup C\ ,\ B\cup C\ ,\ A\cup B\cup C$$

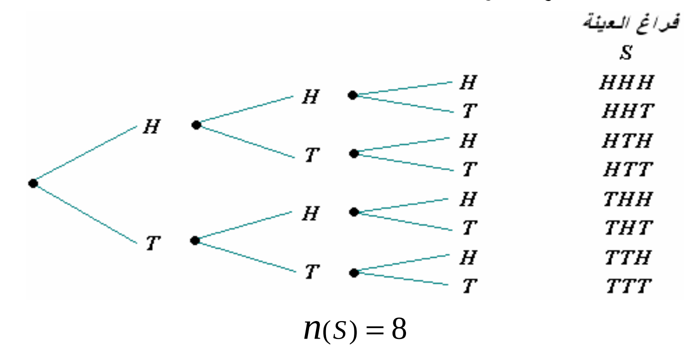
2- إيجاد الأحداث الخاصة بالتقاطعات:

$$A\cap B\ ,\ A\cap C\ ,\ B\cap C\ ,\ A\cap B\cap C$$

3- أوجد الحادث $\overline{B}$

الحـــــــل

- فراغ العينة لهذه التجربة هو:



$$n(S)=8$$

- وأما الأحداث هي:

$A:\{HHT,HTH,THH\}$, $B:\{HTT,THT,TTH\}$, $C:\{HHH,HHT,HTH,HTT\}$

$n(A)=3$ $\quad$ $n(B)=3$ $\quad$ $n(C)=4$

1- الأحداث الخاصة بالاتحاد:

$(A\cup B):\{HHT,HTH,THH,HTT,THT,TTH\}\ ,\ n(A\cup B)=6$

$(A\cup C):\{HHT,HTH,THH,HHH,HTT\}\ ,\ n(A\cup C)=5$

$(B\cup C):\{HHH,HHT,HTH,HTT,THT,TTH\}\ ,\ n(B\cup C)=6$

$(A\cup B\cup C):\{HHH,HHT,HTH,HTT,THT,TTH,THH\}, n(A\cup B\cup C)=7$

2- الأحداث الخاصة بالتقاطع:

$(A\cap B):f\ ,\ n(A\cap B)=0$

$(A\cap C):\{HHT,HTH\}\ ,\ n(A\cap C)=2$

$(B\cap C):\{HTT\}\ ,\ n(B\cap C)=1$

$(A\cap B\cap C):f\ ,\ n(A\cap B\cap C)=0$

3- إيجاد $\overline{B}$:

$(\overline{B}):\{HHH,HHT,HTH,THH,TTT\}\ ,\ n(\overline{B})=5$

3/7 طرق حساب الاحتمالات

يعتمد حساب الاحتمال من الناحية النظرية على أسس وقواعد الرياضيات، ويعتبر هذا النوع من الاحتمال هو العنصر الأساسي في الاستدلال الإحصائي، ولكن في المجال التجريبي تعتمد الاحتمالات على النتائج الفعلية لمشاهدات التجربة، وعلى تكرار الحادث محل الاهتمام، فإذا رمزنا لاحتمال وقوع الحادث A بالرمز $P(A)$، فإن طريقة حساب هذا الاحتمال تتحدد وفقا لنوع الاحتمال، وهما نوعان:

- **الاحتمال التجريبي** Empirical probability: ويعبر عنه بالتكرار النسبي، ويحسب بتطبيق المعادلة التالية:

$$P(A)=\frac{f(A)}{n} \qquad (١-٧)$$

حيث أن: n هو مجموع التكرارات(العدد الكلي للمشاهدات)، $f(A)$: هو تكرار الحادث A،

فإذا تم إلقاء قطعة عملة غير متحيزة 500 مرة، وتم ملاحظة عدد مرات ظهور كل وجه، ولخصت كالتالي:

الوجه (Face)	H	T	SUM
عدد مرات ظهور الوجه	260	240	500

وإذا كان المطلوب حساب احتمال ظهور الصورة H، يمكن تطبيق المعادلة رقم **(7-1)**، والتي تعتمد على التكرار النسبي، أي أن :

$$P(H)=\frac{f(H)}{n}=\frac{260}{500}=0.52$$

- **الاحتمال النظري** Theoretical Probability: وهو الذي يعتمد في حسابه على أسس وقواعد الرياضيات، والتي تستخدم في تحديد عدد النتائج الممكنة للتجربة، وعدد النتائج الممكنة لوقوع الحادث، ومن ثم يحسب هذا النوع من الاحتمال ، بتطبيق المعادلة التالية:

$$P(A)=\frac{n(A)}{n(S)} \qquad (٢-٧)$$

حيث أن: $n(S)$ هو عدد النتائج الممكنة للتجربة، $n(A)$ هو عدد النتائج الممكنة لوقوع الحادث A، فعند إلقاء قطعة عملة غير متحيزة مرة واحدة ، نجد أن فراغ العينة هو: $S:\{H, T\}$ ، أي أن عدد النتائج الممكنة هي: $n(S)=(2)^1 = 2$، وإذا كان الحادث A هو ظهور صورة ، نجد أن $A:\{H\}$ ، أي أن عدد النتائج المكونة للحادث A هي: $n(A)=1$، ويكون احتمال وقوع

الحادث A هو:

$$P(A) = \frac{n(A)}{n(S)} = \frac{1}{2} = 0.5$$

- **العلاقة بين الاحتمال التجريبي و الاحتمال النظري:** عند زيادة عدد المحاولات n يقترب الاحتمال التجريبي من الاحتمال النظري، أي أن:

$$\lim_{n \to \infty} \frac{f(A)}{n} = \frac{n(A)}{n(S)} \qquad (٣-٧)$$

فعند زيادة عدد مرات رمي قطعة العملة، فإن التكرار النسبي للصورة سوف يقترب من القيمة (0.5)، وهي قيمة الاحتمال النظري لظهور الصورة عند رمي قطعة العملة مرة واحدة.

- **النتائج المتشابهة:** إذا أجريت تجربة، وكانت كل نتيجة من النتائج الممكنة للتجربة لها نفس الفرصة في الظهور، بمعنى أن كل نتيجة لها احتمال هو $(1/n(S))$، تسمى هذه النتائج بالنتائج المتماثلة أو المتشابهة، فعند إلقاء زهرة نرد متزنة مرة واحدة، نجد أن فراغ العينة هو S:{1,2,3,4,5,6} ، واحتمال كل نتيجة هو (1/6) ، وعند إلقاء الزهرة مرتين نجد أن عدد نتائج فراغ العينة هو: $n(S)=6^2=36$ نتيجة، وهي:

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

وهذه النتائج متماثلة، واحتمال كل نتيجة هو (1/36).

- **النتائج غير المتماثلة:** هي النتائج التي تحدث عند تكرار محاولة، بحيث أن احتمالات نتائج كل محاولة غير متساوي، ومن ثم لا تتساوى احتمالات نتائج التجربة، فعند سحب كرتين مع الإرجاع بطريقة عشوائية من كيس به ثلاث كرات حمراء (R)، وكرتان تحملان اللون الأبيض (W)، نجد أنه في كل سحب يكون احتمال ظهور كرة حمراء هو 3/5 ، واحتمال ظهور كرة بيضاء هو 2/5 ، ومن ثم يكون نتائج فراغ العينة، واحتمال كل نتيجة في حالة سحب كرتين هو:



يلاحظ أن احتمال كل نتيجة يختلف عن $(1/4)$ ، فهذه الحالات غير متزنة.

4/7 بعض قوانين الاحتمالات Probability Laws

هناك بعض القوانين التي يمكن تطبيقها لحساب الاحتمالات المختلفة، وهي:

- **قانون جمع الاحتمالات** Addition Law

إذا كان لدينا الحادثان A , B ، فإن الاحتمال $P(A\cup B)$ ، يمكن استنتاج معادلته كما يلي:

$$\begin{aligned} P(A\cup B) &= \frac{n(A\cup B)}{n(S)} \\ &= \frac{n(A)+n(B)-n(A\cap B)}{n(S)} \\ &= \frac{n(A)}{n(S)}+\frac{n(B)}{n(S)}-\frac{n(A\cap B)}{n(S)} \\ &= P(A)+P(B)-P(A\cap B) \end{aligned}$$

إذا:

$$P(A\cup B) = P(A)+P(B)-P(A\cap B) \qquad (٧-٣)$$

وفي حالة ثلاث أحداث A , B , C ، يمكن استنتاج معادلة الاتحاد $P(A\cup B\cup C)$ ، وهي:

$$\begin{aligned} P(A\cup B\cup C) &= P(A)+P(B)+P(C) \\ &\quad -P(A\cap B)-P(A\cap C)-P(B\cap C)+P(A\cap B\cap C) \end{aligned} \qquad (٧-٤)$$

وعندما تكون الأحداث متنافية، فإن احتمالات التقاطعات تساوي أصفار، ويكون:

$$P(A \cup B) = P(A) + P(B) \quad ,$$
$$P(A \cup B \cup C) = P(A) + P(B) + P(C) \qquad (٥-٧)$$

مثـــــال(7-2)

عند إلقاء زهرة نرد غير متحيزة مرتين، فأوجد ما يلي:

1- احتمال ظهور وجهين متشابهين.

2- احتمال ظهور وجهين مجموع نقاطهما 10.

3- احتمال ظهور وجهين متشابهين أو مجموع نقاطهما 10.

4- احتمال ظهور وجهين مجموع نقاطهما 7 أو 10.

الحـــــــــل:

نتائج فراغ العينة هي:

S

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

$n(S)=36$

1- بفرض أن الحادث A هو حادث ظهور وجهين متشابهين، فإن:

$A:\{(1,1)\ (2,2)\ (3,3)\ (4,4)\ (5,5)\ (6.6)\},\ n(A)=6$

ويكون احتمال ظهور وجهين متشابهين هو:

$$P(A) = \frac{n(A)}{n(S)} = \frac{6}{36} = \frac{1}{6}$$

2- بفرض أن الحادث B هو حادث ظهور وجهين مجموع نقاطهما 10، فإن:

$B:\{(4,6)\ (5,5)\ (6,4)\},\ n(B)=3$

ويكون احتمال ظهور وجهين متشابهين هو :

$$P(B) = \frac{n(B)}{n(S)} = \frac{3}{36} = \frac{1}{12}$$

3- لحساب احتمال ظهور وجهين متشابهين أو (or) مجموع نقاطهما 10 ، تستخدم المعادلة (7-3)، حيث أن:

$$P(A) = \frac{1}{6} \quad , \quad P(B) = \frac{1}{12}$$

وأما التقاطع $(A \cap B)$ فيعبر عن ظهور وجهين متشابهين و مجموعهما 10 يمكن حسابه كما يلي:

$$(A\cap B):\{(5,5)\}\ ,\ \ n(A\cap B)=1$$

$$P(A\cap B)=\frac{n(A\cap B)}{n(S)}=\frac{1}{36}$$

ومن ثم:

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

$$=\frac{1}{6}+\frac{1}{12}-\frac{1}{36}=\frac{8}{36}=\frac{2}{9}$$

4- بفرض أن الحادث C هو حادث ظهور وجهين مجموع نقاطهما 7، والحادث B هو حادث ظهور وجهين مجموع نقاطهما 10 ، نجد أن:

$B:\{(4,6)\ (5,5)\ (6,4)\}$, $C:\{(1,6)\ (2,5)\ (3,4)\ (4,3)\ (5,2)\ (6,1)\}$

$n(B)=3$ $\qquad$ $n(C)=6$

$P(B)=3/36$ ، $P(C)=6/36$

يلاحظ أن الحادثين C, B حادثين متنافيين، لذا تستخدم المعادلة **(7-5)** في حساب الاحتمال المطلوب كما يلي:

$$P(B\cup C)=P(B)+P(C)=\frac{3}{36}+\frac{6}{36}$$

$$=\frac{9}{36}=\frac{1}{4}$$

- **قانون الاحتمال الشرطي** Conditional probability

يستند هذا الاحتمال على فرصة وقوع حادث، إذا توافرت معلومات عن وقوع حادث آخر له علاقة بالحادث الأول، كاحتمال نجاح الطالب في مادة الإحصاء إذا علم أنه من الناجحين في مادة الاقتصاد، وكاحتمال استخدام المزرعة لنوع معين من السماد، إذا علم أنه يقوم بزراعة محصول معين، وكاحتمال أن الخريجي يعمل بالقطاع الخاص، إذا علم أنه ممن تخرجوا من قسم معين من أقسام كلية الزراعة، والأمثلة على ذلك كثيرة.

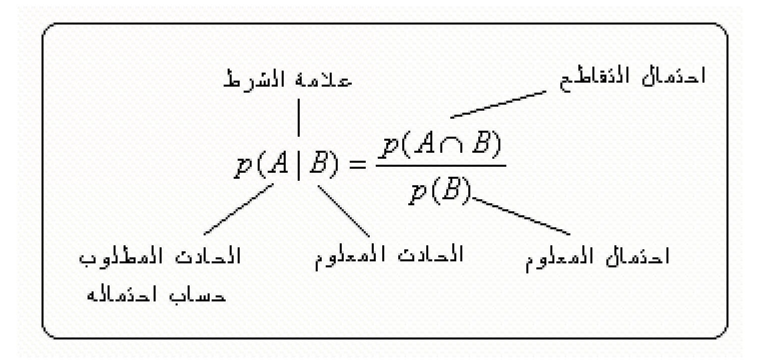
فإذا كان الحادث B حادث معلوم، والحادث A حادث آخر يراد حساب احتمال وقوعه، بمعلومية الحادث B ، فإن هذا الاحتمال يحسب بتطبيق المعادلة التالية:

$$p(A\mid B)=\frac{p(A\cap B)}{p(B)} \qquad (٦-٧)$$

ويعرف الاحتمال $p(A\mid B)$ بقانون الاحتمال الشرطي، ويقرأ "احتمال وقوع الحادث A بمعلومية الحادث B "، أو يقرأ "احتمال وقوع الحادث A بشرط وقوع الحادث B "، كما يمكن حساب احتمال وقوع الحادث B بمعلومية الحادث A ، وذلك بتطبيق المعادلة التالية:

$$p(B\mid A)=\frac{p(A\cap B)}{p(A)} \qquad (٧-٧)$$

ومن المعادلة (7-6)، (7-7) يلاحظ أن الاحتمال الشرطي هو نسبة حادث التقاطع بين إلى الحادث المعلوم، حيث أن :



مثـــــال (7-3)

فيما يلي توزيع تكراري لعينة عشوائية حجمها 100 من خريجي الكلية في العامين الماضيين، حسب التخصص، ونوع المهنة:

التخصص \ المهنة	عمل حكومي	قطاع خاص	عمل حر	Sum
اقتصاد زراعي	15	5	10	30
علوم أغذية	8	17	10	35
علوم تربة	12	10	13	35
Sum	35	32	33	100

فإذا اختير أحد الخريجين بطريقة عشوائية، احسب الاحتمالات التالية:

1- ما احتمال أن يكون من خريجي قسم الاقتصاد و يعمل بالقطاع الخاص.

2- ما احتمال أن يكون ممن يعملون بالحكومة أو من خريجي قسم علوم الأغذية.

3- ما احتمال أن يكون من خريجي قسم علوم الأغذية أو من قسم علوم التربة.

4- إذا علم أن الفرد من خريجي قسم عوم الأغذية، ما احتمال أن يكون ممن يعملون عملا حرا.

الحل:

أولا: نرمز لنوع المهنة بالرمو A، ولنوع التخصص بالرمز B ، كما هو مبين بالجدول التالي:

التخصص \ المهنة	عمل حكومي A_1	قطاع خاص A_2	عمل حر A_3	*Sum*
اقتصاد زراعي B_1	15	5	10	30
علوم أغذية B_2	8	17	10	35
علوم تربة B_3	12	10	13	35
Sum	35	32	33	100

ثانيا: التكرار في كل خلية يعبر عن عدد الخريجين الذين ينتمون لقسم معين و يعملون في مهنة معينة، أي يعبر عن عدد تكرارات حوادث التقاطع الممكنة $A\cap B$.

1- حساب احتمال أن يكون من خريجي قسم الاقتصاد و يعمل بالقطاع الخاص.

$$P(B_1 \cap A_2) = \frac{f(B_1 \cap A_2)}{n} = \frac{5}{100} = 0.05$$

2- حساب احتمال أن يكون ممن يعملون بالحكومة أو من خريجي قسم علوم الأغذية.

$$P(A_1 \cup B_2) = p(A_1) + P(B_2) - P(A_1 \cap B_2)$$
$$= \frac{35}{100} + \frac{35}{100} - \frac{8}{100} = \frac{62}{100} = 0.62$$

3- حساب احتمال أن يكون من خريجي قسم علوم الأغذية أو من قسم علوم التربة.

هذان حادثان متنافيان، لأن تخرج الفرد من أحد الأقسام ينفي تخرجه من الأقسام الآخرى، وبمعنى آخر استحالة أن الفرد تخرج من قسمين في آن واحد، لذا يكون احتمال اتحادهما هو:

$$P(B_2 \cup B_3) = p(B_2) + P(B_3)$$
$$= \frac{35}{100} + \frac{35}{100} = \frac{70}{100} = 0.70$$

4- إذا علم أن الفرد من خريجي قسم عوم الأغذية، ما احتمال أن يكون ممن يعملون عملا حرا، هذا احتمال شرطي، المطلوب هنا " حساب احتمال أن الفرد ممن يعملون عملا حرا A_3 بشرط أنه من خريجي قسم علوم أغذية B_2 ، أي أن الاحتمال المطلوب هو:

$$p(A_3 \mid B_2) = \frac{p(A_3 \cap B_2)}{p(B_2)} = \frac{\left(\frac{10}{100}\right)}{\left(\frac{35}{100}\right)} = \frac{10}{35}$$

واجب منزلي:

الجدول التالي يبين عدد الوحدات السليمة، والتالفة من الخبز العربي بعد ثلاث أيام من تاريخ الإنتاج في أحد مراكز التموين التي تتعامل مع ثلاث مخابز هي : **(C , B , A)** .

الإجمالي	عدد الوحدات التالفة	عدد الوحدات السليمة	
60	24	36	مخبز A
123	63	60	مخبز B
87	33	54	مخبز C
270	120	150	الإجمالي

إذا اختيرت وحدة من الخبز بطريقة عشوائية، فأوجد الآتي:

1- ما احتمال أن تكون من إنتاج المخبز B ؟

2- ما احتمال أن تكون تالفة ؟

3- إذا كانت الوحدة سليمة ، ما احتمال أن تكون من إنتاج المخبز C ؟

4- ما احتمال أن تكون الوحدة من إنتاج المخبز A أو تكون تالفة ؟

5- إذا كانت الوحدة من إنتاج المخبز A، ما احتمال أن تكون تالفة ؟

- **قانون ضرب الاحتمالات** Probability Multiplying Law

ويعكس هذا القانون احتمال وقوع الأحداث معا، أي احتمال التقاطعات، فإذا كـــان A , B ، حادثان يمكن وقوعهما معا، فإن الاحتمال $P(A \cap B)$ يمكن حسابه كحاصل ضرب احتمالين، هما:

$$P(A \cap B) = P(B)\ P(A|B) \quad \text{or} \quad P(A \cap B) = P(A)\ P(B|A) \qquad (٨-٧)$$

مثـــــال (7-4)

إذا كانت نسبة مزارع الخضروات التي تستخدم أسلوب معين للتسميد 60%، وإذا كـــان نسبة المبيعات من إنتاج الخضروات المسمد 70%، بينما نسبة المبيعات من الخضروات غـــير المـــسمدة 80%، إذا اختيرت أحد المزارع التي تنتج الخضروات عشوائيا ، فأوجد الآتي:

1- ما احتمال أن هذه المزرعة تستخدم أسلوب التسميد؟

2- إذا علم أن هذه المزرعة تستخدم أسلوب التسميد، ما احتمال أن تبيع إنتاجها؟

3- ما احتمال أن هذه المزرعة تستخدم أسلوب التسميد وتبيع إنتاجها؟

4- ما احتمال أن هذه المزرعة ممن لا يستخدمون أسلوب التسميد و تبيع إنتاجها؟

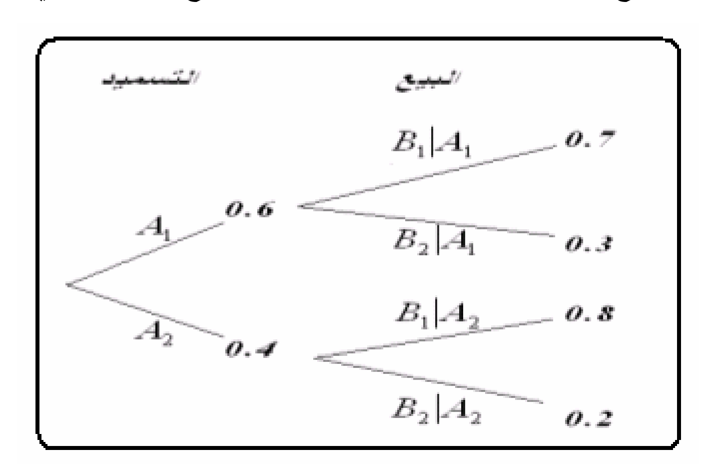
الحــــــل

إذا فحصنا حال المزرعة المسحوبة، نجد أننا نتعامل مع نتيجتين متعاقبتين هما:

النتيجة الأولي ولها حالتان: {المزرعة تستخدم طريقة التسميد (A_1) أو المزرعة لا تستخدم (A_2) }

النتيجة الثانية ولها حالتان: { المزرعة تبيع الإنتاج (B_1)، أو المزرعة لا تبيع الإنتاج (B_2) }

لذا يمكن استنتاج شجرة الاحتمالات للحصول على النتائج الكلية كالتالي:



وفيما يلي حساب الاحتمالات:

1- احتمال أن المزرعة تستخدم أسلوب التسميد هو:

$$P(A_1) = 0.6$$

2- إذا علم أن هذه المزرعة تستخدم أسلوب التسميد، فإن احتمال أن تبيع إنتاجهاهو:

$$P(B_1|A_1) = 0.7$$

3- احتمال أن هذه المزرعة تستخدم أسلوب التسميد وتبيع إنتاجها عبارة عن احتمال وقوع حادثتان معا $(B_1 \text{ and } A_1)$، لذا يحسب هذا الاحتمال بتطبيق المعادلة **(7-8)** كما يلي:

$$P(A_1 \cap B_1) = P(A_1)\ \ P(B_1|A_1)$$
$$= (0.6)(0.7) = 0.42$$

4- احتمال أن المزرعة لا تستخدم أسلوب التسميد وتبيع إنتاجها هو:

$$P(A_2 \cap B_1) = P(A_2)\ \ P(B_1|A_2)$$
$$= (0.4)(0.8) = 0.32$$

- **الأحداث المستقلة** Independent Events

إذا كانت الحادثتان A , B يمكن وقوعهما معا، ولكن وقوع أحدهما ليس له علاقة بوقوع أو عدم وقوع الحادث الآخر، فإن الاحتمال $P(A \cap B)$ يمكن التعبير عنه كالتالي:

$$\boxed{P(A \cap B) = P(A)\ \ P(B)} \qquad (٧-٩)$$

وفي هذه الحالة يقال أن الحاثتان A , B مستقلتان.

مثـــــــال(7-5)

إذا كان نسبة المزارع التي تنتج خضروات 60% ، ونسبة المزارع التي تنتج فاكهــه 75%، ونسبة المزارع التي تنتج الخضروات و الفاكهة 50%، أوجد الآتي:

1- ما احتمال أن مزرعة ما تنتج فاكهة أو خضروات؟

2- ما احتمال ألا تنتج المزرعة الفاكهة ؟

3- هل انتاج المزرعة للفاكهة مستقل عن إنتاجها للخضروات؟

الحــــل:

بفرض أن A حادث يعبر عن "المزرعة تنتج خضروات "، B هو حادث يعبر عن " المزرعة تنتج فاكهة"، فإن:

$$P(A) = 0.6 \ \ , \ \ P(B) = 0.75 \ \ , \ \ P(A \cap B) = 0.5$$

ويكون:

1- احتمال أن مزرعة ما تنتج فاكهة أو خضروات هو:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= (0.6) + (0.75) - 0.5 = 0.85$$

2- احتمال ألا تنتج المزرعة الفاكهة هو:

$$P(\overline{B}) = 1 - P(B) = 1 - 0.75 = 0.25$$

3- لمعرفة ما إذا كان إنتاج المزرعة للفاكهة مستقل عن إنتاجها للخضروات يمكن تطبيق المعادلة (7-9)

$$P(A \cap B) = 0.5 \ , \ P(A)\ P(B) = (0.6)(0.75) = 0.45$$

وحيث أن : $P(A \cap B) \neq P(A)\ P(B)$ **، فإن إنتاج المزرعة للفاكهة** (A)**، غير مستقل عن إنتاجها للخضروات** (B)**.**

مثـــــال(7-6)

إذا كان الحادثان A , B **حادثان مستقلان ، وكان** $P(A) = 0.6$, $P(B) = 0.5$ **، فأوجد الاحتمال** $P(A \cup B)$ **.**

الحـــل:

بما أن الحادثان A ,B **مستقلان، إذا:**

$$P(A \cap B) = P(A)\ P(B)$$
$$= (0.6)(0.5) = 0.3$$

ويكون احتمال $P(A \cup B)$ **هو:**

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= 0.6 + 0.5 - 0.3 = 0.8$$

الفصـــــــل الثامن

المتغيرات العشوائية والتوزيعات الاحتمالية

Random Variables and Probability Distributions

1/8مقـــــدمة

يهتم هذا الفصل بدراسة المتغيرات العشوائية، من حيث تعريفها، وأنواعهـا، والتوزيعـات الاحتمالية لها، وخصائص هذه التوزيعات، والتوزيعات الاحتمالية للمتغيرات العشوائية الخاصة.

2/8 المتغير العشوائي Random Variable:

المتغير العشوائي هو الذي يأخذ قيما حقيقية مختلفة تعبر عن نتائج فراغ العينة، ومن ثم مجـال هذا المتغير، يشمل كل القيم الممكنة له، ويكون لكل قيمة من القيم التي يأخذها المتغير احتمال معين، وينقسم المتغير العشوائي إلى قسمين هما:

1- المتغيرات العشوائية المنفصلة Discrete Random Variables

2- المتغيرات العشوائية المتصلة(المستمرة) Continuous Random Variables

3/8 المتغيرات العشوائية المنفصلة

المتغير العشوائي المنفصل هو الذي يأخذ قيم بينية، ومتباعدة، ويرمز للمتغير العشوائي بشكل عام بحرف من الحروف الأبجدية الكبيرة.... $X, Y, Z,$ ويرمز للقيم التي يأخذها المتغير بـالحروف الأبجديـة الصغيرة، $x, y, z, \dots$، ومن أمثلة هذه المتغيرات:

1- عدد الأولاد الذكور في الأسرة المكونة من أربع أولاد X، $X:\{x=0,1,2,3,4\}$.

2- عدد العملاء الذين يتم إنهاء خدمتهم البنكية كل 10 دقائق Y، $Y:\{y=0,1,2,3,\dots\}$.

3- عدد مرات استخدام نوع معين من الأسمدة خلال الدورة الزراعية.

4- عدد الوحدات التالفة من إنتاج مزرعة معينة تنتج 200 وحدة كل موسم.

5- عدد الوحدات التي تستهلكها الأسرة من سلعة معينة خلال الشهر.

وهكذا..... الأمثلة كثيرة

1/3/8 التوزيع الاحتمالي للمتغير العشوائي المنفصل

التوزيع الاحتمالي، هو الذي يبين احتمالات حدوث القيم التي يمكن يأخذها المتغير، والتي تـرتبط باحتمالات النتائج الممكنة في فراغ العينة، وبمعنى آخر هو التكراري النسبي للقيم التي يمكن أن يأخذها المتغير.

فإذا كان المتغير العشوائي المنفـصل X يأخـذ القـيم، $X:\{x=x_1,x_2,\dots,x_n\}$ ، وكـان

$P(X = x_i) = f(x_i)$ هو احتمال أن المتغير العشوائي يأخذ القيمة x_i ، فإنه، يمكـــن تكـــوين جدول التوزيع الاحتمالي للمتغير العشوائي X ، وهو جدول مكون من عمودين، الأول به القـــيم الممكنة للمـــتغير $X : \{x = x_1, x_2, ..., x_n\}$ ، والثـــاني بـــه القـــيم الاحتماليـــة لهـــذا المـــتغير $P(X = x_i) = f(x_i)$ ، أي أن:

جدول **(8-1)**

جدول التوزيع الاحتمالي للمتغير العشوائي المنفصل

x_i	$f(x_i)$
x_1	$f(x_1)$
x_2	$f(x_2)$
$\vdots$	$\vdots$
x_n	$f(x_n)$
Σ	1

وتسمى الدالة $f(x_i)$ بدالة الاحتمال، ومن خصائص هذه الدالة ما يلي:

$$\begin{array}{ll} 1- & 0 < f(x_i) < 1 \\ 2- & \sum f(x_i) = 1 \end{array} \qquad (١-٨)$$

مثـــــال **(8-1)**

إذا كان من المعلوم أن نسبة مبيعات أحد المراكز التجارية من التفاح الأمريكي 0.60 ، بينما يكون نسبة مبيعاته من الأنواع الأخرى للتفاح 0.40، اشترى أحد العملاء عبوتين، والمطلوب:

1- كون فراغ العينة.

2- إذا عرف المتغير العشوائي X بأنه عدد العبوات المشتراة من التفاح الأمريكي، فأوجد الآتي:

- التوزيع الاحتمالي للمتغير العشوائي X .
- ارسم دالة الاحتمال لهذا المتغير.
- كون التوزيع الاحتمالي التجميعي.
- ما هو احتمال $P(X = 1)$ ، $P(X \leq 1)$ ، $P(X = 1.5)$ ، $P(X \leq 1.5)$
- حدد قيمة الوسيط، والمنوال لعدد العبوات المشتراة.

الحـــــل:

تكوين فراغ العينة:

التجربة هنا هو شراء وحدتين من عبوات التفاح، ومن ثم فراغ العينة يتكون من أربع نتائج، هي:

		S	عدد العبوات X	$P(X=x)=f(x)$
أمريكي 0.60	أمريكي 0.60	(أمريكي،أمريكي)	2	0.36
	آخر 0.40	(آخر,أمريكي)	1	0.24
آخر 0.40	أمريكي 0.60	(أمريكي,آخر)	1	0.24
	آخر 0.40	(آخر,آخر)	0	0.16

- **التوزيع الاحتمالي لعدد العبوات المشتراة من التفاح الأمريكي** X

من المعلوم أن العميل اشترى عبوتين، وأن المتغير العشوائي هو عدد العبوات المشتراة من التفـاح الأمريكي، لذا تكون القيم الممكنة للمتغير العشوائي هي:

$x=0$ إذا كانت العبوتين من النوع الآخر، أى إذا كانت نتيجة التجربة (آخر، آخر)

$x=1$ إذا كان أحد العبوتين من النوع الأمريكي، أي إذا كانت نتيجة التجربة (آخر، أمريكـي) أو (أمريكي،آخر)

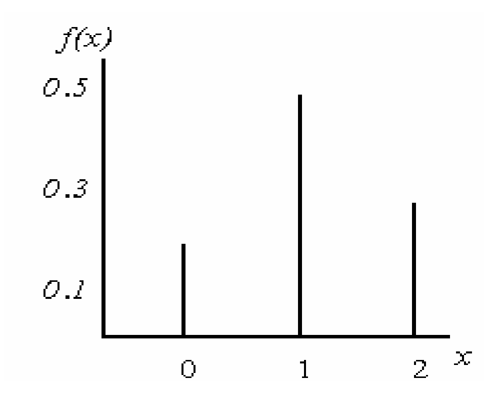
$x=2$ إذا كان العبوتين من النوع الأمريكي، أي إذا كانت نتيجة التجربة (أمريكي ، أمريكي)

ومن ثم يأخذ المتغير القيم: $X:\{x=0,1,2\}$ ، ويرتبط احتمالات هذه القيم باحتمالات نتائج التجربة المناظرة لها كما هو مبين أعلاه، ومن ثم يكون التوزيع الاحتمالي للمتغير العشوائي X هو:

جدول التوزيع الاحتمالي لعدد العبوات المشتراة من التفاح الأمريكي

x_i	$f(x_i)$
0	0.16
1	0.48
2	0.36
Σ	1

- **رسم دالة الاحتمال $f(x)$:**



- **تكوين التوزيع الاحتمالي التجميعي:**

التوزيع التجميعي، هو جدول يشمل الاحتمالات الناتجة من حساب الاحتمال $P(X \leq x)$، ويرمز

له بالرمز $F(x)$ ، أي أن دالة التوزيع الاحتمالي التجميعي تأخذ الصورة التالية:

$$F(x) = P(X \le x) \quad (٨-٢)$$

ومن ثم يمكن تكوين جدول التوزيع الاحتمالي التجميعي لعدد الوحدات المشتراة من التفاح الأمريكي كما يلي:

جدول التوزيع الاحتمالي، والتوزيع التجميعي لعدد العبوات المشتراه من التفاح الأمريكي

x_i	$f(x_i)$	$F(x_i)$
0	0.16	$F(0) = P(X \le 0) = 0.16$
1	0.48	$F(1) = P(X \le 1) = 0.16 + 0.48 = 0.64$
2	0.36	$F(2) = P(X \le 2) = 0.64 + 0.36 = 1.00$
Σ	1	

- **حساب الاحتمالات:-** $P(X = 1)$ ، $P(X \le 1)$ ، $P(X = 1.5)$ ، $P(X \le 1.5)$

$$P(X = 1) = f(1) = 0.48$$
$$P(X \le 1) = F(1) = 0.64$$
$$P(X = 1.5) = f(1.5) = 0$$
$$P(X \le 1.5) = F(1.5) = F(1) = 0.64$$

- **تحديد قيمة الوسيط، والمنوال.**

الوسيط:- رتبة الوسيط هو 0.50 ، إذا الوسيط M هو القيمة التي تحقق الاحتمال: $P(X \le M) = F(M) = 0.50$ ، وهذا الاحتمال يقع بين القيمتين *(1,0)* كما هو مبين بالرسم التالي:

	x_i	$F(x_i)$	
M	0	0.16	
			$F(M) = 0.50$
	1	0.64	
	2	1.00	

إذا الوسيط قيمته هي:

$$M = 0 + \frac{0.5 - 0.16}{0.64 - 0.16} \times (1 - 0) = 0.71$$

حساب المنوال:

المنوال $Mode$ = القيمة x_i المناظرة لأكبر قيمة احتمالية.

إذا المنوال هو: $Mode = 1$ حيث أنه يناظر أكبر قيمة احتمالية هي: $f(1) = 0.48$.

2/3/8 الوسط الحسابي والتباين للمتغير العشوائي المنفصل

أ- يرمز للوسط الحسابي للمتغير العشوائي بالرمز m (ميو)، ويحسب بتطبيق المعادلة التالية:

$$\mu = \sum x_i \ f(x_i) \qquad (٨-٣)$$

ب- وأما التباين ويرمز له بالرمز s^2 (سيجما)، فيحسب بتطبيق المعادلة التالية:

$$\begin{aligned}\sigma^2 &= \sum (x_i - \mu)^2 \ f(x_i) \\ &= \sum x_i^2 \ f(x_i) - \mu^2\end{aligned} \qquad (٨-٤)$$

مثـــال (8-2)

في المثال السابق احسب ما يلي:

أ- الوسط الحسابي لعدد العبوات المشتراة من النوع الأمريكي:

ب- احسب الانحراف المعياري لعدد العبوات المشتراة من النوع الأمريكي.

ت- أوجد معامل الاختلاف النسبي:

الحـــل

أ- الوسط الحسابي لعدد العبوات من النوع الأمريكي:

لحساب الوسط الحسابي والانحراف المعياري يتم استخدام المعادلة (8-3)، (8-4) وهذا يتطلــب تكوين جدول يشمل المجاميع التالية: $\sum x_i \ f(x_i) \ , \ \sum x_i^2 \ f(x_i)$، وذلك كما يلي:

x_i	$f(x_i)$	$x_i f(x_i)$	$x_i^2 f(x_i)$
0	0.16	0	0
1	0.48	0.48	0.48
2	0.36	0.72	1.44
Σ	1	1.20	1.92

إذا الوسط الحسابي هو: $m = \sum x_i \ f(x_i) = 1.20$

ب- ولحساب الانحراف المعياري يجب أولا حساب التباين وهو:

$$s^2 = \sum x_i^2 \ f(x_i) - m^2 = 1.92 - (1.20)^2 = 0.48$$

إذا الانحراف المعياري قيمته هي:

$$s = \sqrt{s^2} = \sqrt{0.48} = 0.693$$

ت- معامل الاختلاف النسبي هو:

$$C.V=\frac{s}{m}\times 100=\frac{0.693}{1.2}\times 100=57.7$$

واجب منزلي:-

فيما يلي التوزيع الاحتمالي لعدد الوحدات التي تستهلكها الأسرة من أحد مساحيق النظافة خلال الشهر X ، $X:\{x=0,1,2,3,4,5\}$

x (عدد الوحدات التي تستهلكها الأسرة)	0	1	2	3	4	5
$f(x)$	0.15	0.30	0.25	0.23	0.05	0.02

والمطلوب:

1- حدد نوع هذا المتغير (عدد الوحدات التي تستهلكها الأسرة)
2- احسب الوسط والوسيط والمنوال والانحراف المعياري لعدد الوحدات المستهلكة.
3- كون جدول التوزيع التجميعي $F(x)$ ثم أوجد الآتي:
أ- نسبة الأسر التي يقل استهلاكها عن وحدتين
ب- نسبة الأسر التي يزيد استهلاكها عن 3 وحدات
ت- إذا كان لدينا 500 أسرة، فما هو عدد الأسر المتوقع أن يكون استهلاكها على الأقل 3 وحدات؟
4- احسب معامل الالتواء، وكذلك معامل الاختلاف النسبي، وعلق على النتائج.

4/8 التوزيعات الاحتمالية المنفصلة الخاصة

في كثير من النواحي التطبيقية، تتبع بعض الظواهر توزيعات احتمالية خاصة، وهي التوزيعات التي يمكن حساب احتمالات قيم المتغير عن طريق معادلة رياضية، تسمى بدالة الاحتمال $f(x)$ ، وهذه المعادلة لها معالم معينة، تسمى بمعالم المجتمع الذي ينسب له هذا التوزيع، وهذه المعالم ما هي إلا حقائق ثابتة مجهولة، وهي الأساس في حساب القيم الاحتمالية للتوزيع الاحتمالي للمجتمع محل الدراسة.

ومن أهم التوزيعات التي سيتم دراستها في هذا المقرر، توزيع ثنائي الحدين، والتوزيع البواسون.

1/4/8 التوزيع ثنائي الحدين The Binomial Distribution

يستخدم هذا التوزيع في الحالات التي يكون للظاهرة محل الدراسة نتيجتان فقط متنافيتان، النتيجة محل الاهتمام وتسمى بحالة النجاح، والأخرى تسمى بحالة الفشل، ومن أمثلة ذلك:

- **عند إعطاء مريض نوع معين من الأدوية، لها نتيجتان: (استجابة للدواء، أو عدم استجابة)**
- **عند فحص عبوة بداخلها نوع معين من الفاكهة، لها نتيجتان (الوحدة إما أن تكون سليمة، أو تكون معيبة)**
- **عند إلقاء قطعة عملة، لها نتيجتان (ظهور الوجه الذي يحمل الصورة، أو الوجه الذي يحمل الكتابة)**
- **نتيجة الطالب في الاختبار (نجاح، رسوب)**

- استخدام المزارع لبرنامج معين في الزراعة (يستخدم، أو لا يستخدم).

شكل التوزيع الاحتمالي ثنائي الحدين

إذا كررت محاولة n من المرات، بحيث أن كل محاولة لها نتيجتان فقط متنافيتان هما:

- النتيجة محل الاهتمام " حالة نجاح " وتتم باحتمال ثابت في كل محاولة هو p
- النتيجة الأخرى " حالة فشل " وتتم باحتمال ثابت أيضا هو $q = 1 - p$

وبافتراض أن هذه المحاولات مستقلة، بمعنى أن نتيجة كل محاولة ليس لها علاقة بنتيجة المحاولة الأخرى، وإذا كان المتغير العشوائي X يعبر عن عدد حالات النجاح "عدد النتائج محل الاهتمام" في الـ n محاولة، فإن مدي المتغير العشوائي X والذي يعبر عن عدد حالات النجاح هو: $X : \{x = 0,1,2,...,n\}$ ، ومن ثم يحسب الاحتمال $P(X = x) = f(x)$ بتطبيق المعادلة التالية:

$$f(x) = \binom{n}{x} p^x q^{n-x} \quad , \quad x = 0,1,2,...n \qquad (٨-٥)$$

حيث أن $\binom{n}{x}$ هي عدد طرق اختيار x من n مع إهمال الترتيب، وتحسب كما يلي:

$$\binom{n}{x} = \frac{n(n-1)(n-2)...(n-x+1)}{x(x-1)(x-2)...3\times2\times1} \qquad (٨-٦)$$

$$\binom{7}{3} = \frac{7\times6\times5}{3\times2\times1} = 35 \;=\; \binom{7}{4}$$

$$\binom{7}{0} = \binom{7}{7} = 1$$

مثـــــال (8-3)

إذا كان من المعلوم أن نسبة الشفاء من مرض معين باستخدام نوع معين من العقاقير الطبية هو 0.60 ، إذا تناول هذا العقار 5 مصابين بهذا المرض. إذا عرف المتغير العشوائي X بأنه عدد الـذين المستجيبين (حالات الشفاء) لهذا العقار.

المطلوب:

أ- ما هو نوع المتغير؟

ب- اكتب شكل دالة الاحتمال $f(x)$ لهذا المتغير.

ت- احسب الاحتمالات التالية:

- ما احتمال استجابة 3 مرضى لهذا العقار؟
- ما هو احتمال استجابة مريض واحد على الأقل؟
- ما هو احتمال استجابة 2 مرضى على الأكثر؟

ث- احسب الوسط الحسابي، والانحراف المعياري لعدد حالات الاستجابة.

ج- حدد شكل التوزيع.

الحـــــل:

أ- عدد حالات الاستجابة X متغير كمي منفصل ، ومدى هذا المتغير في هذه الحالة هـو:
$X : \{x = 0,1,2,3,4,5\}$:

ب- شكل دالة الاحتمال: $n = 5$ ، $p = 0.60$ ، $q = 1 - p = 0.40$ إذا:

$$f(x) = \binom{n}{x}(p)^x (q)^{n-x}$$

$$= \binom{5}{x}(0.6)^x (0.4)^{5-x} \quad , \quad x = 0,1,2,3,4,5$$

ت- حساب الاحتمالات:

- حساب احتمال استجابة 3 مرضى لهذا الدواء: $P(x=3) = f(3)$

$$f(3) = \binom{5}{3}(0.6)^3 (0.4)^{5-3} = \frac{5\times4\times3}{3\times2\times1}\times0.216\times0.16 = 10\times0.03456$$
$$= 0.3456$$

- حساب احتمال استجابة مريض واحد على الأقل: $P(x \ge 1)$

$$P(x \ge 1) = f(1) + f(2) + f(3) + f(4) + f(5) = 1 - f(0)$$
$$= 1 - \left[\binom{5}{0}(0.6)^0 (0.4)^5\right] = 1 - 1\times1\times0.01024 = 0.98976$$

- حساب احتمال استجابة 2 مرضى على الأكثر: $P(x \le 2)$:

$$P(x \le 2) = f(2) + f(1) + f(0)$$

$$= \binom{5}{2}(0.6)^2 (0.4)^3 + \binom{5}{1}(0.6)^1 (0.4)^4 + \binom{5}{0}(0.6)^0 (0.4)^5$$
$$= \frac{5\times4}{2\times1}(0.36)(0.064) + \frac{5}{1}(0.6)(0.0256) + 1(1)(0.01024$$
$$= 0.2304 + 0.0768 + 0.01024 = 0.31744$$

ث- حساب الوسط الحسابي، والانحراف المعياري لعدد حالات الاستجابة:

- الوسط الحسابي (m) في حالة التوزيع ثنائي الحدين يحسب بتطبيق المعادلة **(8-3)**، وباستخدام العمليات الرياضية يمكن الوصول إلى النتيجة التالية:

$$\mu = \sum x\ f(x) = np \qquad (٨-٧)$$

إذا الوسط الحسابي هو:

$$m = np = 5(0.60) = 3$$

- الانحراف المعياري هو الجذر التربيعي الموجب للتباين، ولحساب التباين في التوزيع ثنائي الحدين يتم تطبيق المعادلة **(8-4)**، ومنها يمكن التوصل إلى

الصورة التالية:

$$\sigma^2 = npq \qquad (٨-٨)$$

إذا تباين عدد حالات الاستجابة هو:

$$s^2 = npq$$
$$= 5(0.60)(0.40) = 1.2$$

ومن ثم يأخذ الانحراف المعياري الصورة التالية:

$$s = \sqrt{npq}$$
$$= \sqrt{1.2} = 1.095$$

ويمكن حساب معامل الاختلاف النسبي، بتطبيق المعادلة التالية:

$$V.C = \frac{s}{m} \times 100 = \frac{1.095}{3} \times 100 = 36.5\%$$

ج- تحديد شكل التوزيع:

يتحدد شكل التوزيع ثنائي الحدين وفقا لقيمة احتمال النجاح p كما يلي:

إذا كان $p = 0.5$ فإن التوزيع الاحتمالي ثنائي الحدين يكون متماثل.

إذا كان $p < 0.5$ فإن التوزيع الاحتمالي ثنائي الحدين يكون موجب الالتواء.

إذا كان $p > 0.5$ فإن التوزيع الاحتمالي ثنائي الحدين يكون سالب الالتواء.

وحيث أن $p = 0.6 > 0.5$ فإن توزيع عدد حالات الاستجابة سالب الالتواء.

2/4/8 التوزيع البواسوني Poisson Distribution

يكثر استخدام هذا التوزيع في الحالات التي تقع فيها الأحداث وفقا لمعدلات زمنية، وكذلك في حالة الأحداث نادرة الوقوع، ومن أمثلة ذلك:

- عدد الوحدات التي تستهلكها الأسرة من سلعة معينة خلال الشهر. $X : \{x = 0,1,2,...\}$
- عدد مرات ري نوع معين من المحاصيل الزراعية خلال الموسم. $X : \{x = 0,1,2,...\}$
- عدد العملاء الذين يتم خدمتهم البنكية كل 10 دقائق. $X : \{x = 0,1,2,...\}$
- عدد مرات زيارة المريض للطبيب كل سنة. $X : \{x = 0,1,2,...\}$
- عدد مرات تناول الأسرة للحوم الحمراء خلال الأسبوع. $X : \{x = 0,1,2,...\}$
- عدد أخطاء الطباعة لكل صفحة من صفحات الكتاب. $X : \{x = 0,1,2,...\}$

وهكذا الأمثلة كثيرة

شكل التوزيع الاحتمالي البواسوني

إذا كان متوسط عدد مرات وقوع حادث وفقا لمعدل زمني معين هو m، وكان المتغير العشوائي X يعبر عن عدد مرات وقوع الحادث وفقا لهذا المعدل، فإن مدي المتغير العشوائي X هو: $X : \{x = 0,1,2,...\}$، وهذا المدى عبارة عن فئة مفتوحة من اليمين، فإن الاحتمال

$P(X=x)=f(x)$ والذي يعبر عن احتمال وقوع الحادث عدد x من المرات وفقا لهـــذا المعـــدل، يحسب بتطبيق المعادلة التالية:

$$f(x)=\frac{e^{-\mu}\mu^{x}}{x!} \quad , \quad x=0,1,2,... \qquad (٨-٩)$$

حيث أن e هي أساس اللوغاريتم الطبيعي، وتوجد في بعض الآلات الحاسبة، وقيمتها هي: $e=2.718$ تقريبا، ويمكن حساب قيمتها باستخدام الآلة الحاسبة باتباع الخطوات التالية من الشمال إلى اليمين:

مثلا إيجاد $e^{-1.5}$

النتيجة — (0.22323016) (=) (1.5) ((-)) (e^x) (SHIFT)

وأما $x!$ فتسمى "مضروب العدد x " ويساوي: $x!=x(x-1)(x-2)...3\times2\times1$

مثـــــال (8-4)

إذا كان من المعلوم أن عدد الوحدات التي تستهلكها الأسرة من سلعة معينة خلال الشهر تتبع توزيع بواسون بمتوسط 3 وحدات شهريا، إذا عرف المتغير العشوائي X بأنه عدد الوحـــدات الـــتي تستهلكها الأسرة خلال الشهر من هذه السلعة.

المطلوب:

أ- ما هو نوع المتغير العشوائي؟

ب- اكتب شكل دالة الاحتمال $f(x)$ لهذا المتغير.

ح- احسب الاحتمالات التالية:

- احتمال أن الأسرة تستهلك وحدتين خلال الشهر؟
- احتمال أن أسرة ما تستهلك وحدة واحد على الأقل خلال الشهر؟
- احتمال أن أسرة ما تستهلك 3 وحدات على الأكثر خلال الشهر؟

خ- احسب الوسط الحسابي، والانحراف المعياري لعدد الوحدات المستهلكة.

د- حدد شكل التوزيع.

الحـــل:

أ- عدد الوحدات التي تستهلكها الأسرة X متغير كمي منفصل ، ومدى هذا المتغير في هذه الحالة هو: $X:\{x=0,1,2,3,...\}$:

ب- شكل دالة الاحتمال:

بما أن متوسط عدد الوحدات التي تستهلكها الأسرة خلال الشهر هو: $m=3$، إذا دالة الاحتمال هي:

$$f(x)=\frac{e^{-m}m^{x}}{x!}$$

$$=\frac{e^{-3}3^{x}}{x!}\quad , \quad x=0,1,2,\ldots$$

ح- حساب الاحتمالات:

- حساب احتمال أن أسرة ما تستهلك وحدتين خلال الشهر، *f(2)*

$$f(2)=\frac{e^{-3}3^{2}}{2!}=\frac{0.0498(9)}{2\times1}=0.22404$$

- احتمال أن أسرة ما تستهلك وحدة واحد على الأقل خلال الشهر هو:

$$P(X\geq1)=f(1)+f(2)+\ldots.$$

$$=1-f(0)=1-\frac{e^{-3}3^{0}}{0!}=\frac{0.0498}{1}=1-0.0498=0.9502$$

- احتمال أن أسرة ما تستهلك 3 وحدات على الأكثر خلال الشهر هو:

$$P(X\leq3)=f(3)+f(2)+f(1)+f(0)$$

$$=\frac{e^{-3}3^{3}}{3!}+\frac{e^{-3}3^{2}}{2!}+\frac{e^{-3}3^{1}}{1!}+\frac{e^{-3}3^{0}}{0!}\frac{0.0498}{1}$$

$$=0.0498\left(\frac{27}{6}+\frac{9}{2}+\frac{3}{1}+\frac{1}{1}\right)=0.0498(13)=0.6474$$

خ- حساب الوسط الحسابي، والانحراف المعياري لعدد حالات الاستجابة:

- الوسط الحسابي (m) في حالة التوزيع البواسون هو معلمة معطاة هي:

$$m=3$$

في هذا التوزيع، فإن التباين يساوي الوسط الحسابي:

أي أن: $s^{2}=m=3$

ومن ثم يكون الانحراف المعياري هو:

$$s=\sqrt{m}=\sqrt{3}=1.732$$

ويمكن حساب معامل الاختلاف النسبي، بتطبيق المعادلة التي سبق استخدامها في الفصل السابق، وهو:

$$V.C=\frac{s}{m}\times100=\frac{1.732}{3}\times100=57.7\%$$

د- تحديد شكل التوزيع:

دائما التوزيع البواسون موجب الالتواء.

5/8 المتغيرات العشوائية المستمرة Continuous Random Variables

المتغير العشوائي المستمر، هو الذي يأخذ قيما متصلة، ويأخذ عدد لانهائي من القيم الممكنة له داخـــل مجالـــه، فـــإذا كـــان X مـــتغير عـــشوائي مـــستمر، ويقـــع في المـــدى (a,b)، أي أن: $\{X = x : a < x < b\}$، فإن للمتغير X عدد لانهائي من القيم تقع بين الحـــدين الأدنى والأعلـــى (a,b)، ومن الأمثلة على المتغيرات الكمية المستمرة ما يلي:

- كمية الألبان التي تنتجها البقرة في اليوم باللتر: $\{X = x : 10 < x < 40\}$
- المساحة المتزرعة بالأعلاف في المملكة بالألف هكتار $\{X = x : 1000 < x < 15000\}$
- فترة صلاحية حفظ الدجاج المبرد بالأيام، $\{X = x : 1 < x < 5\}$
- وزن الجسم بالكيلوجرام للأعمار من *(30-40)*، $\{X = x : 55 < x < 80\}$

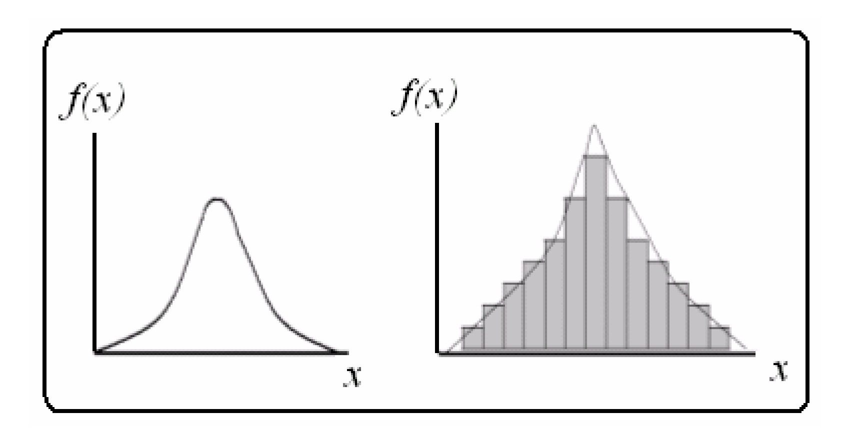
وهكذا الأمثلة على المتغير الكمي المستمر كثيرة.

1/5/8 التوزيع الاحتمالي للمتغير المستمر Continuous Probability

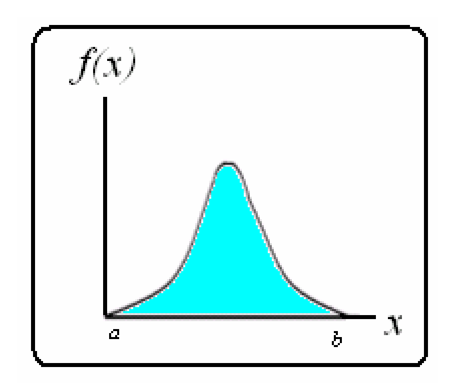
عند تمثيل بيانات المتغير الكمي المستمر في شكل مدرج تكراري النسبي، نجد أن شكل هذا المدرج هو أقرب وصف لمنحنى التوزيع الاحتمالي للمتغير المستمر، وكلما ضاقت الفترات بين مراكز الفئات، يمكن الحصول على رسم دقيق للمنحنى الخاص بدالة احتمال المتغير المستمر، كما هو مـــبين بالـــشكل التالي:

شكل **(8-1)**

شكل منحنى التوزيع الاحتمالي للمتغير العشوائي المستمر



والمساحة أسفل المنحنى تعبر عن مجموع الاحتمالات الكلية، ولذا تساوي هذه المساحة الواحد الصحيح، وتسمى الدالة $f(x)$ بدالة كثافة الاحتمال Probability Distribution Function$(p.d.f)$ ، وبفرض المتغير العشوائي المستمر يقع في المدى: $X=\{x: a \pi x \pi b\}$ ، وأن منحنى هذه الدالة يأخذ الصورة التالية:



فإن من خصائص دالة كثافة الاحتمال $f(x)$ ما يلي:

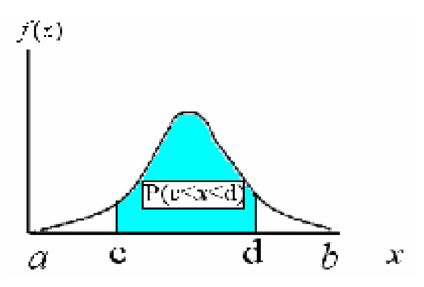
1- الدالة $f(x)$ موجبة داخل المدى (a,b) أي أن: $f(x) \phi 0$ ، $x \in (a,b)$

2- التكامل على حدود المتغير من الحد الأدنى a حتى الحد الأعلى b يعبر عن مجموع الاحتمالات الكلية، لذا يساوي الواحد الصحيح ، أي أن:

$$\int_{x=a}^{x=b} f(x)\ dx = 1 \qquad (١٠-٨)$$

حيث أن الشكل الرياضي أعلاه يسمى بالتكامل المحدد من $x=a$ حتى $x=b$ ، وهذا يعني إيجاد المساحة أسفل المنحني بين (a,b) .

3- لحساب احتمال أن المتغير العشوائي المستمر يقع في المدى (d,c) أي حساب الاحتمال $p(c<x<d)$ ، يجب حساب المساحة أسفل المنحني من $x=c$ حتى $x=d$ كما هي مبينة في الشكل البياني التالي:



ويتم ذلك بإيجاد التكامل المحدد في هذا المدى، كما يلي:

$$p(c<x<d)=\int_{x=c}^{x=d} f(x)\,dx=[g(x)]_c^d=g(d)-g(c) \qquad (١١-٨)$$

4- في المتغير المستمر، يكون الاحتمال $p(x=value)$ مساويا للصفر، أي أن:

$$p(x=value)=0 \qquad (١١-٨)$$

ولكي يمكننا حساب الاحتمالات، يجب عرض بعض قواعد التكامل التالية:

جدول (8-2)

بعض قواعد التكامل

(1)	$\int x^n dx=\frac{x^{n+1}}{n+1}$ and $\int (a+bx)^n dx=\frac{(a+bx)^{n+1}}{b(n+1)}$ (2) $\int e^x dx=e^x$ and $\int e^{(a+bx)}dx=\frac{1}{b}e^{(a+bx)}$ (3) $\int \frac{1}{x}dx=\log_e(x)$ and $\int \frac{1}{(a+bx)}dx=\frac{1}{b}\log_{e(a+bx)}$	integration
(4)	$\Gamma(n+1)=\int_0^\infty x^n e^{-x}dx=n!=n(n-1)(n-2)...3\times2\times1$	gamma
(5)	$I\Gamma(n+1)=\int_0^a x^n e^{-x}dx=n!\left(1-e^{-a}\sum_{i=0}^{n}\frac{a^i}{i!}\right)$	Incomplete gamma
(6)	$B(m+1,n+1)=\int_0^1 x^n(1-x)^m dx=\frac{m!n!}{(m+n+1)!}$	Beta

مثـــال (8-5)

إذا كان الإنفاق الشهري للأسرة بالألف ريال على المواد الغذائية له دالة كثافة احتمال تأخذ الصورة التالية:

$$f(x)=\begin{cases} cx(10-x) & , 0<x<10 \\ 0 & otherwise \end{cases}$$

والمطلوب:

1- حساب قيمة الثابت c

2- احسب احتمال أن إنفاق الأسرة يتراوح ما بين (8,5) ألف ريال خلال الشهر.

3- إذا كان لدينا 600 أسرة، فما هو عدد الأسر المتوقع أن يقل إنفاقها عـــن 3 آلاف خـــلال الشهر؟

الحـــــــــل

1- حساب قيمة c

من خصائص دالة كثافة الاحتمال:

$$\int_{x=a}^{x=b} f(x)\ dx = 1$$

إذا

$$\int_{x=0}^{x=10} cx(10-x)\ dx = c\int_{x=0}^{x=10}(10x-x^2)\ dx = c\left[10\left(\frac{x^2}{2}\right)-\frac{x^3}{3}\right]_0^{10}$$

$$= c\left[5x^2-\frac{x^3}{3}\right]_0^{10} = c\left[(5(100)-\frac{(1000)}{3})\right]-0$$

$$= \frac{500}{3}c = 1$$

$$c = 3/500 = 0.006$$

2- حساب أن إنفاق الأسرة يتراوح بين (8,5) ألف ريال خلا الشهر هو.

$$p(5<x<8) = \int_{x=5}^{x=8} 0.006x(10-x)\ dx = 0.006\left[5x^2-\frac{x^3}{3}\right]_5^8$$

$$= 0.006\left[\left(5(8)^2-\frac{8^3}{3}\right)-\left(5(5)^2-\frac{5^3}{3}\right)\right] = 0.006[(149.3333)-(83.3333)]$$

$$= 0.006(66) = 0.396$$

3- إذا كان لدينا 600 أسرة، فإن عدد الأسر المتوقع أن يقل إنفاقها عن 3 آلاف خلال الـــشهر هو:

$$number\ of\ \ family = 600\ \ p(x<3)$$

$$= 600\int_0^3 0.006x(10-x)dx$$

$$= 3.6\left[5x^2-\frac{x^3}{3}\right]_0^3 = 3.6[45-9]-0 = 129.6 \approx 130$$

حوالي 130 أسرة.

2/5/8 المتوسط والتباين في التوزيع الاحتمالي المستمر

إذا كانت $f(x)$ هي دالة كثافة الاحتمال للمتغير العشوائي x، $a<x<b$ فإن التوقع الرياضي للدالة $h(x)$ تأخذ الصورة التالية:

$$E(h(x)) = \int_a^b h(x)\ dx \qquad (١٢-٨)$$

ومن ثم يمكن كتابة معادلة الوسط والتباين كما يلي.

$$\mu = E(x) = \int_a^b xf(x)dx \qquad (١٣-٨)$$

$$\sigma^2 = E(x^2) - u^2\ ,\ E(x^2) = \int_a^b x^2 f(x)\ dx$$

تابع مثال (8-5)

في المثال السابق أوجد المتوسط والانحراف المعياري ومعامل الاختلاف النسبي للإنفاق الشهري.

الحـــــل

1- المتوسط الحسابي

$$m = E(x) = xf(x)dx = \int_0^{10} x(0.006x(10-x)) = 0.006\int_0^{10}(10x^2 - x^3)dx$$

$$= 0.006\left[10\frac{x^3}{3} - \frac{x^4}{4}\right]_0^{10} = 0.006\left[\left(\frac{10000}{3} - \frac{10000}{4}\right) - (0)\right]$$

$$= 60\left[\frac{1}{12}\right] = 5$$

متوسط إنفاق الأسرة الشهري 5 آلاف ريال.

2- الانحراف المعياري

$$s^2 = E(x^2) - u^2 = E(x^2) - (5)^2$$

$$E(x^2) = \int_a^b x^2 f(x)\ dx = 0.006\int_0^{10}(10x^3 - x^4)dx$$

$$= 0.006\left[10\left(\frac{x^4}{4}\right) - \left(\frac{x^5}{5}\right)\right]_0^{10} = 0.006\left[\frac{100000}{4} - \frac{100000}{5}\right] - 0$$

$$= 600\left(\frac{1}{20}\right) = 30$$

إذا التباين هو : $s^2 = 30 - 25 = 5$ ، ومن ثم يأخذ الانحراف المعياري القيمة التالية:

$$s = \sqrt{\text{var}\,iance} = \sqrt{5} = 2.236$$

3- معامل الاختلاف النسبي

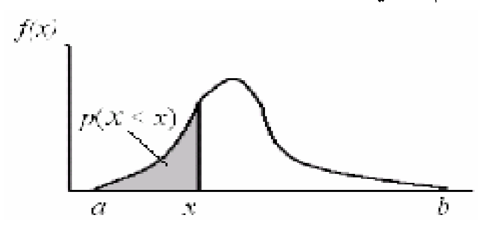
$$C.V = \frac{s}{m} \times 100 = \frac{2.236}{5} \times 100 = 44.72\%$$

دالة التوزيع التجميعي Cumulative Distribution Function (C.D.F)

يرمز لهذه الدالة بالرمز *(C.D.F)=F(x)* وتحسب بإيجاد الاحتمال:

$$C.D.F = F(x) = P(X \leq x) = \int_{a}^{x} f(x)dx \qquad (١٤-٨)$$

ويمكن توضيحها بيانيا بالرسم التالي:



تابع مثال (5-8)

في المثال (**5-8**) أوجد دالة التوزيع التجميعي *C.D.F*، ثم استخدم هذه الدالة لحساب احتمال أن إنفاق الأسرة يقل عن 5 آلاف جنيه.

الحـــــل

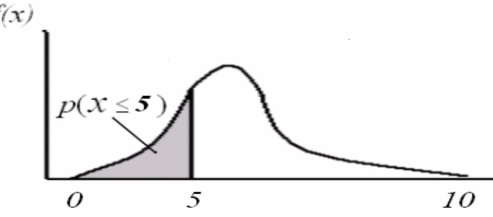
- إيجاد دالة التوزيع التجميعي *C.D.F*

$$F(x) = \int_{0}^{x} f(x)\ dx$$

$$= \int_{0}^{x} 0.006x(10-x)dx = 0.006\left[10\left(\frac{x^2}{2}\right) - \left(\frac{x^3}{3}\right)\right]_0^x$$

$$= 0.006\left[5x^2 - \left(\frac{x^3}{3}\right)\right]$$

- حساب الاحتمال المطلوب $F(5) = p(x \leq 5)$ ، كما هو مبين بالرسم التالي:



ويمكن حساب هذا الاحتمال بالتعويض عن $x = 5$ في الدالة *F(x)* التي تم التوصل إليها، أي أن:

$$F(5) = P(x \le 5) =$$
$$= 0.006\left[5x^2 - \frac{x^3}{3}\right] = 0.006\left[125 - \frac{125}{3}\right]$$
$$= 0.006\left(\frac{250}{3}\right) = 0.5$$

أي أن 50% من الأسر يقل إنفاقها عن 5 آلاف ريال.

خصائص دالة التوزيع التجميعي

1- $F(x) \phi 0$	**2-** $F(a) = 0$	**3-** $F(b) = 1$	**4-** $p(X \phi x) = 1 - F(x)$
5- $f(x) = dF(x)/dx$			

6/8 التوزيعات الاحتمالية المستمرة الخاصة

Continuous Probability Distributions

هناك بعض التوزيعات الاحتمالية المستمرة الخاصة، ولها دوال كثافة احتمال محددة، وفيما يلي بعض هذه التوزيعات:

1/6/8 التوزيع المنتظم Uniform distribution

شكل دالة كثافة الاحتمال *p.d.f*

هو توزيع له دالة احتمال ثابتة، ويستخدم في حالة الظواهر التي يمكن أن تحدث بشكل منتظم، فإذا كان المتغير x متغير عشوائي له توزيع منتظم *Uniform*، مداه هو $a < x < b$ فإن دالة كثافة احتماله هي:

$$f(x) = \frac{1}{b-a}\ ,\ a < x < b \qquad (١٥-٨)$$

ويمكن تمثيل هذه الدالة بيانيا كما يلي:

معالم هذا التوزيع

توجد معلمتان لهذا التوزيع هما (b,a)، ولذا يكتب رمز لهذا التوزيع الصورة $x \sim U(a,b)$

خصائص التوزيع المستطيل

الوسط الحسابي m ، والتباين s^2 لهذا المتغير هما :

$$m = E(x) = \frac{a+b}{2} \ , \quad s^2 = \frac{(b-a)^2}{12}$$

على الطالب إثبات ذلك:

دالة التوزيع التجميعي $C.D.F$

تأخذ دالة التوزيع التجميعي $F(x)$ الشكل الآتي

$$F(x) = p(X \le x) = \int_a^x f(x)dx = \frac{1}{b-a}\int_a^x dx = \frac{x-a}{b-a} \qquad (١٦-٨)$$

مثـــال (8-6)

استورد أحد المراكز التجارية 1500 طن بطاطس، ووضعها في مخزن، وقام ببيعها بكميـــات متساوية على مدار شهور السنة. إذا كانت الفترة الزمنية للبيع تتبع توزيع منتظم، فأوجد الآتي:

- دالة كثافة الاحتمال المعبرة عن الفترة الزمنية للبيع.
- بعد مرور سبعة أشهر من بداية البيع، ما هي الكمية الموجودة بالمخزن؟

الحــــــــل

- دالة كثافة الاحتمال المعبرة عن الزمن:

بفرض أن المتغير x يعبر عن الفترة الزمنية للبيع مقاسة بالـــشهر، أي أن $0 < x < 12$ ، ومن ثم تأخذ دالة كثافة الاحتمال المعبرة عن الزمن الصورة التالية:

$$f(x) = \frac{1}{12-0} = \frac{1}{12} \ , \ 0 < x < 12$$

- حساب الكمية الموجودة بالمخزن بعد سبعة أشهر من بداية البيع.

بفرض أن Q هي كمية البطاطس المستوردة ، تكون الكية المتبقية بالمخزن بعد مـــرور سبعة أشهر من بداية البيع هي :

$$Q \times p(x > 7) = Q \times (1 - F(7)) = 1500(1 - \frac{7-0}{12-0}) = 625 \ Ton$$

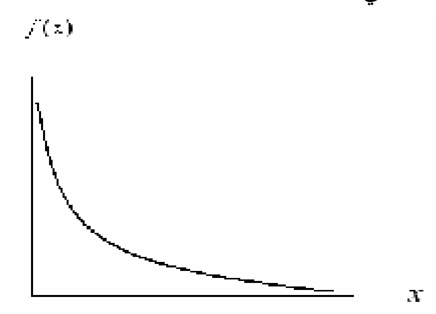
2/6/8 التوزيع الأسي السالب Negative Exponential distribution

شكل دالة كثافة الاحتمال $p.d.f$

إذا كان المتغير x متغير عشوائي له توزيع أسي سالب ، مداه هو $0 < x < \infty$ فإن دالـــة كثافة احتماله هي:

$$f(x) = \theta\, e^{-\theta x},\ 0 < x < \infty\ ,\ \theta \succ 0 \qquad (١٧-٨)$$

ويمكن تمثيل هذه الدالة بيانيا كما يلي:



معالم هذا التوزيع

توجد معلمة واحدة هي (q)

خصائص التوزيع الأسى السالب

الوسط الحسابي m ، والتباين s^2 لهذا المتغير هما:

$$m = E(x) = \frac{1}{q}\ ,\quad s^2 = \frac{1}{q^2}$$

دالة التوزيع التجميعي $C.D.F$

تأخذ دالة التوزيع التجميعي $F(x)$ الشكل الآتي

$$F(x) = p(X \le x) = \int_0^x f(x)dx = \left(1 - e^{-qx}\right)$$

مثـــــال (8-7)

إذا كانت الفترة الزمنية لإنهاء خدمة العميل في البنك تتبع توزيع أسي بمتوسط 2 دقيقة، فأوجد ما يلي.

- دالة كثافة الاحتمال المعبرة عن الفترة الزمنية لإنهاء خدمة العميل.
- ما احتمال إنهاء خدمة العميل في أقل من دقيقة.

الحـــــل

- دالة كثافة الاحتمال المعبرة عن الزمن:

بفرض أن المتغير x يعبر عن الفترة الزمنية لإنهاء خدمة العميل بالدقيقة، أي أن $0 < x < \infty$ ، فإن المتوسط $1/q = 2$ ، ومن ثم تصبح قيمة (q) هي: $(q = 0.5)$ ، وتكتب دالة كثافة الاحتمال المعبرة عن الزمن على الصورة التالية:

$$f(x) = 0.5\ e^{-0.5\ x},\ 0 < x < \infty$$

- حساب احتمال إنهاء خدمة العميل في أقل من دقيقة.

$$p(x \le 1) = (1 - e^{-0.5x}) = (1 - e^{-0.5(1)}) = 0.3935$$

3/6/8 التوزيع الطبيعي The Normal Distribution

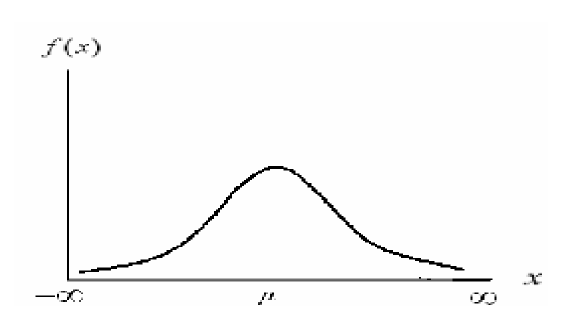
يعتبر هذا التوزيع من أكثر التوزيعات الاحتمالية استخداما في النواحي التطبيقية، ومنها الاستدلال الإحصائي شاملا التقدير، واختبارات الفروض، كما أن معظم التوزيعات يمكن تقريبها إلى هذا التوزيع، وفيما يلي عرض لهذا التوزيع.

شكل دالة كثافة الاحتمال $p.d.f$

إذا كان المتغير x متغير عشوائي له توزيع طبيعي ، مداه هو $-\infty < x < \infty$ فإن دالة كثافة احتماله هي:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},\ -\infty < x < \infty\ ,\ \pi = 22/7 \qquad (١٨-٨)$$

وهذا التوزيع له منحنى متماثل يأخذ الصورة التالية:



فهذا المنحنى متماثل على جانبي الوسط الحسابي m .

معالم هذا التوزيع

توجد معلمتين لهذا التوزيع هما :

الوسط الحسابي : $E(x) = m$ والتباين : $\text{var}(x) = s^2$

ومن ثم يعبر عن توزيع المتغير x بالرموز : $x \sim N(m, s^2)$ ويعني ذلك أن المتغير العشوائي x

يتبع التوزيع الطبيعي بمتوسط m ، وتباين s^2 .

خصائص التوزيع الطبيعي

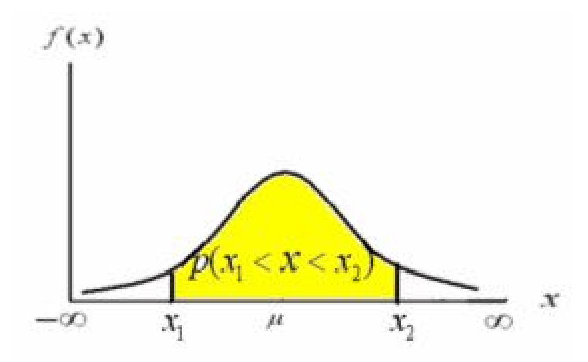
هذا التوزيع من أكثر التوزيعات الاحتمالية استخداما، بل يشتق منه كل التوزيعات الاحتمالية الأخرى المستخدمة في الاستدلال الإحصائي، ومن خصائص هذا التوزيع ما يلي:

1- الوسط الحسابي m **2**- والتباين s^2

3- منحني هذا التوزيع متماثل على جانبي الوسط m

كيفية حساب الاحتمالات $p(x_1 < X < x_2)$

بفرض أن الاحتمال المطلوب حسابه هو $p(x_1 < X < x_2)$ ، وهذا الاحتمال يحدد بالمساحة التالية:



وحيث أن هذا التوزيع من التوزيعات المستمرة، فإن هذه المساحة (الاحتمال) تحسب بإيجاد التكامل التالي:

$$p(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)dx = \int_{x_1}^{x_2} \frac{1}{s\sqrt{2p}} e^{-\frac{1}{2}\left(\frac{x-m}{s}\right)^2} dx$$

وهذا التكامل يصعب حسابه، ومن ثم لجأ الإحصائيين إلى عمل تحويلة رياضية Transform، يمكن استخدام توزيعها الاحتمالي في حساب مثل هذه الاحتمالات، وهذه التحويلة هي:

$$z = \left(\frac{x-m}{s}\right)$$

ويعرف المتغير الجديد z بالمتغير الطبيعي القياسي Standard Normal Variable ، وهذا المتغير له دالة كثافة احتمال تأخذ الصورة التالية:

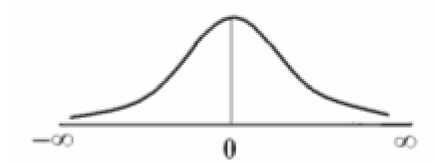
$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} , -\infty < z < \infty \;,\; \pi = 22/7 \qquad (١٩-٨)$$

ومن خصائص هذا التوزيع ما يلي:

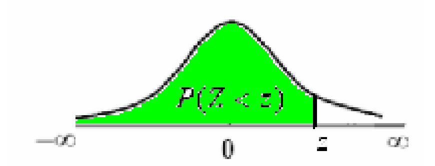
1- متوسطه هو: $E(z) = 0$ **2**- تباينه هو: $\text{var}(z) = 1$

ومن ثم يعبر عن توزيع المتغير z بالرموز : $z \sim N(0,1)$ ويعني ذلك أن المتغير العشوائي x يتبع التوزيع الطبيعي القياسي بمتوسط (0) ، وتباين (1) .

3- يأخذ المنحنى الشكل الناقوس المتماثل على جانبي الصفر:



وصمم الإحصائيون جداول إحصائية لحساب دالة التوزيع التجميعي: $F(z) = P(Z < z)$ ، كما هو مبين بالرسم التالي:

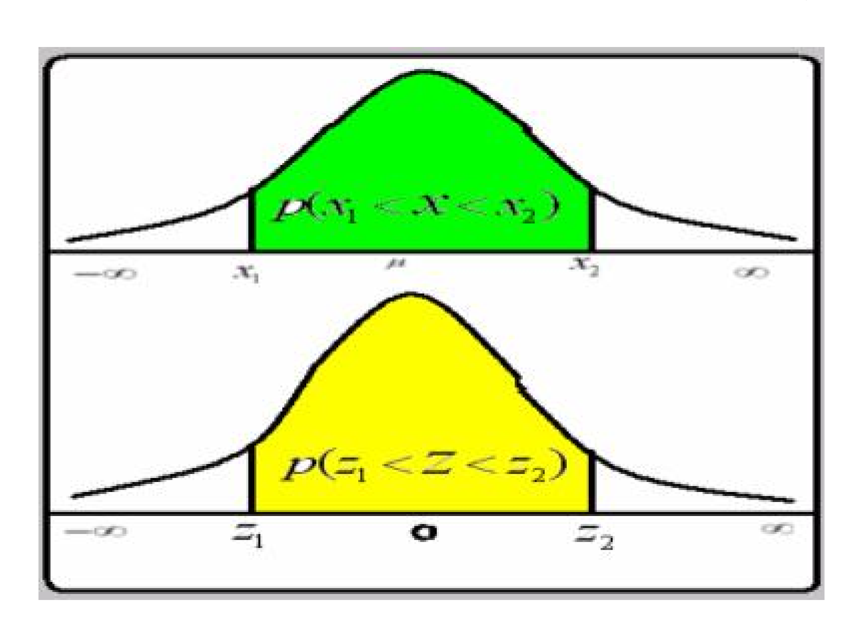


ونعود الآن إلى خطوات حساب الاحتمال $p(x_1 < X < x_2)$ باستخدام التحويلة $z = (x - m)/s$:

1- يتم تحويل القيم الطبيعية (x_1, x_2) إلى قيم طبيعية قياسية:

$z_1 = (x_1 - m)/s \quad , \quad z_2 = (x_2 - m)/s$.

2- ومن ثم يكون الاحتمال: $p(x_1 < X < x_2) = p(z_1 < Z < z_2)$:



3- تستخدم جداول التوزيع الطبيعي القياسي، والــذي يعطــي المســاحة الخاصــة بالاحتمــال $F(z) = P(Z < z)$

4- طريقة استخدام جدول التوزيع الطبيعي القياسي في حساب الاحتمالات

أوجد الاحتمالات التالية:

أ- $P(z < 1.57)$ ب- $P(z < -2.33)$ ج- $P(z > 1.96)$ د-

$P(-2.01 < z < 1.28)$

الحل

أ- تحدد المساحة المعبرة عن الاحتمال $P(z < 1.57) = F(1.57)$ أسفل المنحنى كما يلي



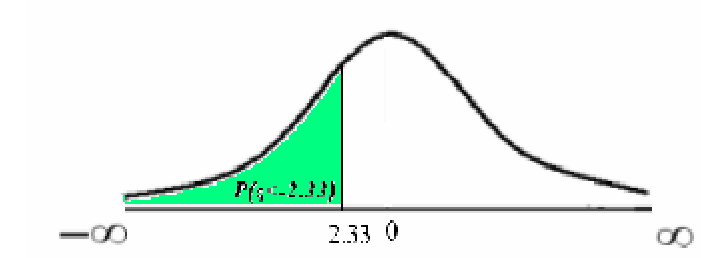
ويتم استخدام الجدول كما هو مبين :

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
⋮										
1.00										
1.10										
1.20										
1.30										
1.40										
1.50								0.9418		
⋮										

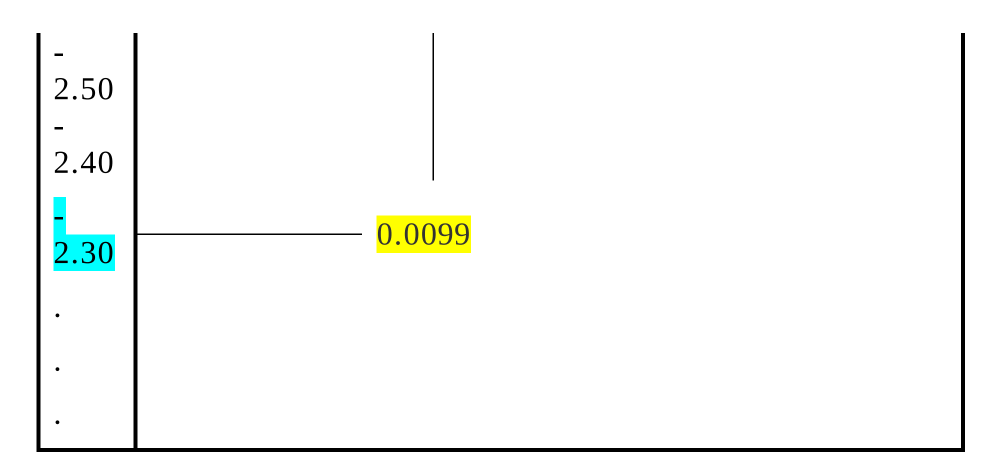
ويكون الاحتمال المطلوب هو: $P(z < 1.57) = F(1.57) = 0.9418$

ب- المساحة أسفل المنحنى المعبرة عن الاحتمـــال $P(z < -2.33) = F(-2.33)$ موضـــحة كالتالي:

P(z<-2.33)



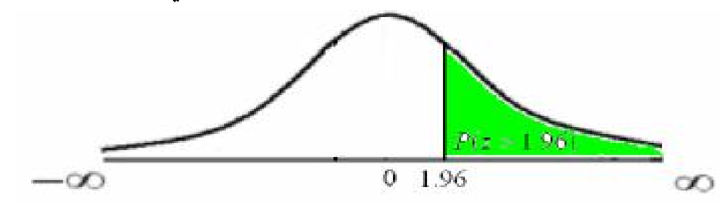
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.										
.										
.										
.										
-2.70										
-2.60										



-2.50	
-2.40	
-2.30	0.0099
.	
.	
.	

ومن ثم يكون : $P(z<-2.33)=0.0099$

ج- تحدد المساحة المعبرة عن الاحتمال $P(z>1.96)$ كالتالي:



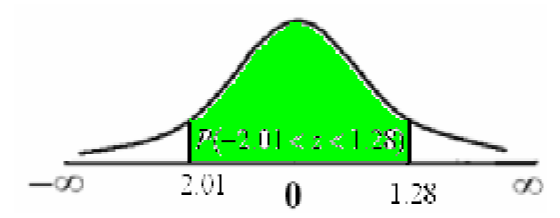
وهذا الاحتمال يحسب باستخدام خصائص دالة التوزيع التجميعي ، حيث أن :

$$P(z>1.96)=1-p(z<1.96)=1-F(1.96)$$

وبالكشف في الجدول بنفس الطريقة السابقة على القيمة 1.96 نجد أن : $p(z<1.96)=0.9750$ ، ومن ثم يكون الاحتمال المطلوب هو:

$$P(z>1.96)=1-0.9750=0.0250$$

د- المساحة أسفل المنحنى المعبرة عن الاحتمال $P(-2.01<z<1.28)$ هي:



وباستخدام أيضا خصائص دالة التوزيع التجميعي يمكن حساب هذا الاحتمال ، حيث أن :

$$P(-2.01<z<1.28)=F(1.28)-F(-2.01)$$

وبالكشف في الجدول عن هاتين القيمتين ، نجد أن:

$$P(-2.01<z<1.28)=0.8997-0.0222=0.8775$$

مثـــــــال(8-8)

إذا كان الدخل السنوي للأسرة في أحد مناطق المملكة يتبع توزيع طبيعي متوسطه 80 ألـف ريال، وتباينه 900. والمطلوب:

1- كتابة قيمة معالم التوزيع الاحتمالي للدخل السنوي.

2- كتابة شكل دالة كثافة الاحتمال.

3- ما هي نسبة الأسر التي يقل دخلها عن60 ألف ريال ؟

4- ما هو الدخل الذي أقل منه 0.975 من الدخول؟

الحـــــــل

1- كتابة قيمة معالم التوزيع الاحتمالي للدخل السنوي.

بفرض أن x متغير عشوائي يعبر عن الدخل السنوي بالألف ريال، وهو يتبع التوزيـــع الطبيعـــي، ومعالمه هي:

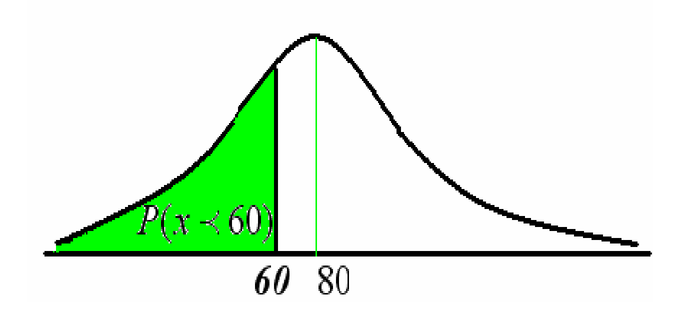
أ- المتوسط $E(x)=m=80$ ب- التباين هو: $Var(x)=s^{2}=900$

أي أن : $x \sim N(80,900)$

2- شكل دالة كثافة الاحتمال

$$f(x)=\frac{1}{30\sqrt{2p}}\; e^{-\frac{1}{2}\left(\frac{x-80}{30}\right)^{2}} \;,\; -\infty < x < \infty \;,\; p = 22/7$$

3- نسبة الأسر التي يقل دخلها عن60 ألف ريال هي: $P(x \pi 60)$



ويتبع الخطوات المذكورة سابقا في حساب الاحتمال كما يلي:

$$P(x<60)=p\left(z<\frac{x-m}{s}\right)$$

$$=P\left(z<\frac{60-80}{30}\right)=P(z<-0.67)=F(-0.67)$$

وبالكشف مباشرة عن هذه القيمة في جدول التوزيع الطبيعي القياسي ، نجد أن

$$P(x<60)=P(z<-0.67)=0.2514$$

4- الدخل الذي أقل منه 0.975 من الدخول: في هذه الحالة يبحث عن قيمة المتغير (x) الذي أقل منه 0.975 ، بفرض أن هذا المتغير هو (x_1) ، فإن :

$$P(x < x_1) = p\left(z < \frac{x_1 - 80}{30}\right) = 0.975$$

بالكشف بطريقة عكسية ، حيث نبحث عن المساحة 0.9750 نجدها تقع عند تقاطع الـــصف 1.9 ، والعمود 06. أي أن قيمة $z = 1.96$ ، ويكون :

$$1.96 = \frac{x_1 - 80}{30} \quad , \quad Then \quad x_1 = 30(1.96) + 80 = 138.8$$

إذا الدخل هو 138.8 ألف ريال في السنة.